Systematic Analysis and Optimization of Broadband Noise and Linearity in Silicon Germanium Heterojunction Bipolar Transistors

A Thesis
Presented to
The Academic Faculty

by

Qingqing Liang

In Partial Fulfillment
of the Requirements for the Degree of
Doctor of Philosophy

School of Electrical and Computer Engineering
Georgia Institute of Technology
December 2004

Systematic Analysis and Optimization of Broadband Noise and Linearity in Silicon Germanium Heterojunction Bipolar Transistors

Approved by:

Dr. John D. Cressler, Advisor
School of Electrical & Computer Engineering
Georgia Institute of Technology

Dr. Xu-Yan Chen
School of Mathemetics
Georgia Institute of Technology

Dr. Joy Laskar
School of Electrical & Computer Engineering
Georgia Institute of Technology

Dr. Mary Ann Ingram
School of Electrical & Computer Engineering
Georgia Institute of Technology

Dr. Marshall Leach
School of Electrical & Computer Engineering
Georgia Institute of Technology

Date Approved:

December 2004

ACKNOWLEDGEMENTS

I would like to take this opportunity to thank all those who have helped me in my research over the years, and those who are so important to me.

I would like to express my deepest thanks to my advisor and co-advisor, Dr. John D. Cressler and Dr. Guofu Niu. For the past four years they have stimulated and guided me as I explored the field of device physics, modeling, analysis, and optimization. They showed me the paradigm of scholars, and taught me the ethics of scientists. They will always remain my respected professors, and I will be grateful to them all my life.

I would like to sincerely thank Dr. Joy Laskar, Dr. Marshall Leach, Dr. Mary Ann Ingram, and Dr. Xu-Yan Chen, who served as my committee members, for all the valuable comments they gave me, which helped me a lot in conducting my research, and thanks for all their patience as they reviewed and corrected my dissertation.

Thanks are due to all the professors who instructed me in their lectures. Their teaching sharpened my skills and made me more proficient. I would also like to thank Ms. Palmer and Ms. Grindel for helping me improve my professional communication skills.

Thank you to all my colleagues in our group: Alvin Joseph, David Sheridan, Shiming Zhang, Gang Zhang, Jongsoo Lee, Emery Chen, Ying Li, Zhenrong Jin, Zhiyun Luo, Tianbing Chen, Ramkumar Krithivasan, Chendong Zhu, Yuan Lu, Jin Tang, Yun Shi, Joel Andrews, Lance Kuo, Jon Comeau, Xiangtao Li, Enhai Zhao, Bhaskar Banerjee, Sunitha Venkataraman, Curtis Grens, Akil Sutton, Becca Haugerud, Adnan Ahmed, Mustayeen Nayeem, Mustansir Pratapgarhwala, and Mingteng Han. Their help has solidified my research work.

I also want to thank my parents, Xuegong Liang and Fan Huang, for their support and enlightenment in my youth. They aroused my interest in science and engineering, and offered me the best education within their abilities. It is they who opened the doors for me.

Finally, I want to thank my wife, Jinghong Ma, for her understanding and encouragements during my graduate studies. She will always be my good companion in my future work and life.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

Noise and linearity are the two key concerns in RF transceiver systems. However, the impact of circuit topology and device technology on systems' noise and linearity behaviors is poorly understood because of the complexity and diversity involved. There are two general questions that are addressed by the RF device and circuit designers: for a given device technology, how best to optimize the circuit topology; and for a given circuit topology, how best to optimize the device technology to improve the noise and linearity performance.

In this dissertation, a systematic noise and linearity calculation method is proposed. This approach offers simple and analytical solutions to optimize the noise and linearity characteristics of integrated circuits. Supported by this approach, the physics of state-of-the-art SiGe HBT technology devices can be decoupled and studied. The corresponding impact on noise and linearity is investigated. New optimization methodologies for noise and linearity at both the device and circuit level are presented.

In addition, this thesis demonstrates a technique that accurately extracts *ac* and noise parameters of devices/circuits in the millimeter-wave range. The extraction technique supports and verifies the device/circuit noise analysis from a measurement standpoint.

CHAPTER I

INTRODUCTION

1.1 The development and application of SiGe HBTs

The theory of hetero-junction transistors (HBTs) and bandgap engineering was pioneered by Kroemer in 1957 [1]. The idea is to improve the gain and speed of bipolar junction transistors using the exponential factors induced by bandgap offsets. However, early manufactured HBTs were all III-V type (e.g., GaAs) and expensive. It was not until the late 1980s that the growth of SiGe on pure Si epitaxy using the ultra-high vacuum/chemical vapor deposition (UHV/CVD) technique [2] was developed, providing an economical approach to building HBTs on silicon. Figure 1 shows a cross-section of a first-generation SiGe HBT manufactured by IBM. The Ge profile was grown in the neutral base region to improve device performance (e.g., high gain, high speed).

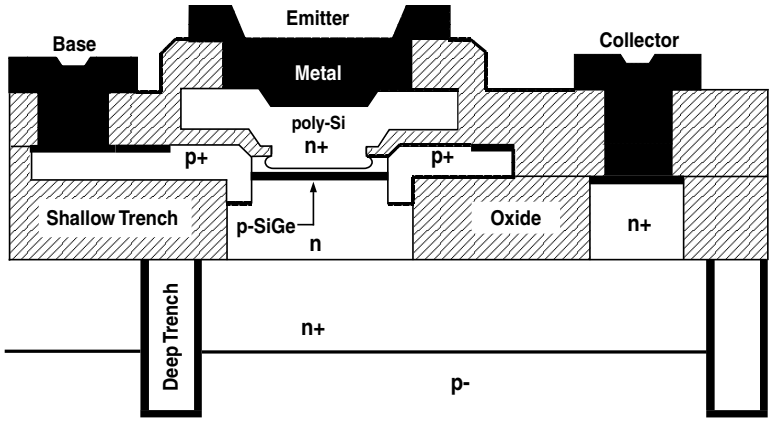

Figure 1: A cross-section of a first-generation SiGe HBT [4].

Since the first functional SiGe HBT was developed in 1987 [3], SiGe HBTs have rapidly emerged as an important alternative to III-V HBTs for radio-frequency (RF) and mixed-signal applications [4] because of their high performance and low cost. The research and development activity in SiGe devices, circuits, and technologies in both industry and at universities worldwide has grown rapidly since then [5]-[7]. The figures-of-merit of SiGe HBTs, correspondingly, continue

to improve dramatically. Figures 2 and 3 show the historical trends in peak cut-off frequency f_T and peak maximum oscillation frequency f_{max} for integrated, self-aligned SiGe HBT and SiGeC HBT technologies, respectively. Note from the curves that in the past two years, the reported peak f_T and f_{max} of SiGe HBTs have improved to above 300 GHz, which makes possible millimeter-wave ($f > 30$ GHz) applications using Si-based devices.

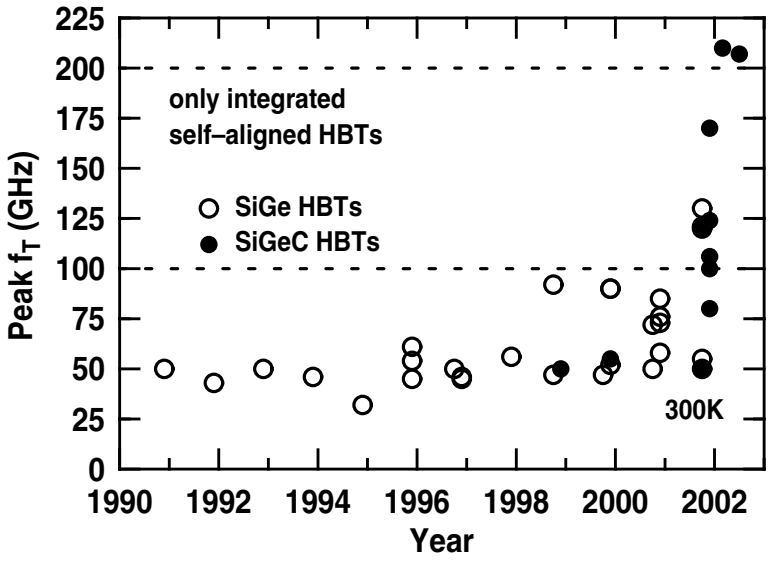

Figure 2: Historical trends in peak cut-off frequency for integrated, self-aligned SiGe HBT and SiGeC HBT technologies [8].

This rapid development of device technology and high-frequency applications has consequently exposed many new challenges in characterizing and modeling SiGe HBTs. Much research has been devoted to examine these device-related problems, such as high-injection barrier effects [9][10], high-frequency substrate coupling [11], self-heating [12][13], non-quasi-static effects [14], etc. However, the impact of these phenomena on noise and linearity is still poorly understood and modeled. For instance, in current compact models, the cut-off frequency f_T roll-off at high injection and the SiGe retrograde barrier effect are lumped into one equation. However, this assumption will be invalid in deep SiGe/Si interface devices, and the consequent effects are thus neglected. On the other hand, a key advantage of SiGe HBTs is their ability to be device-level optimized for a specified circuit design (e.g., a low-noise amplifier, or a power amplifier). This requires the device engineer to understand the impact of these effects on circuit performance. Therefore, a direct link between device physics and circuit performance needs to be established.

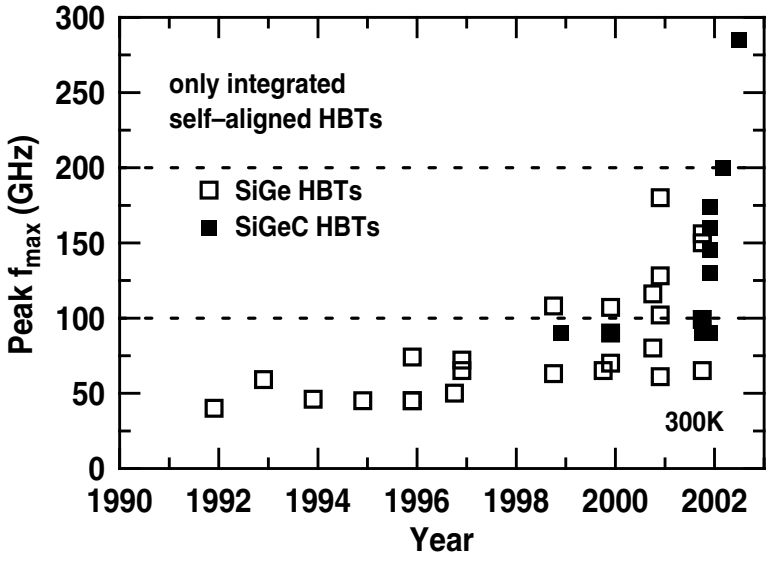

Figure 3: Historical trends in peak maximum oscillation frequency for integrated, self-aligned SiGe HBT and SiGeC HBT technologies [8].

1.2 Broadband noise extraction and modeling

The noise processes of various types are a fundamental aspect of the universe. Thus, noise is far more pervasive than just in the field of electrical engineering, where it is usually formally treated. In this thesis, the noise process in device physics and its impact on integrated circuit are discussed.

The first impedance-field approach to calculating noise in semiconductor devices was presented by Shockley in 1966 [15]. Various noise types in semiconductor devices were then explored and modeled by Van der Ziel in 1968 [16]. Since then, solid-state noise theory has become an established field of study. Recently, high-frequency noise characterization (broadband noise) has presented a new challenge and needs to be re-investigated.

Broadband noise is one of the key considerations of RFIC front-end designs, including low-noise amplifiers (LNAs) [17][18], mixers [19][20], and voltage-control oscillators (VCOs) [21], because it directly determines the lower limit of the input RF signal power level (sensitivity) [22].

Figure 4 illustrates the noise factor N and noise figure NF of a noisy amplifier. The signal-to-noise ratio (SNR) at the output is always lower than the SNR at the input because of the noise generated by the amplifier. Optimization of the device or circuit is required to obtain a better output SNR. Therefore, accurately extracting and modeling device broadband noise are important.

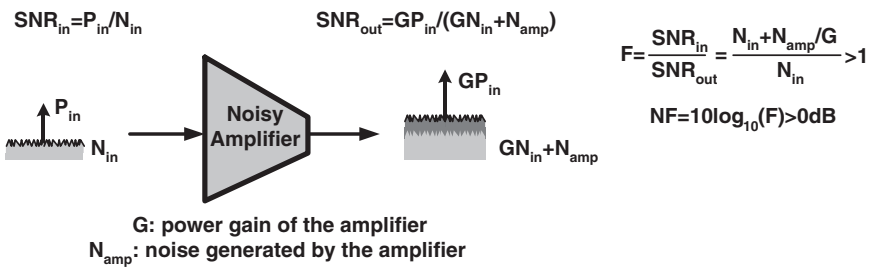

Figure 4: The noise factor F and noise figure NF of a noisy amplifier.

The first problem is noise measurement and extraction. In a two-port system, the broadband noise is characterized by three parameters: minimum noise figure NF_{min}, noise impedance R_n, and optimum conductance Y_{OPT} for noise matching [23]. By measuring these three parameters the noise behavior of the system can be characterized [24].

In high-frequency s-parameter and broadband noise measurements, the on-wafer parasitics induced by the pad capacitance, interconnection impedance, and substrate coupling are significant. Figure 5 shows the distributive on-wafer parasitics of a SiGe device used for RF characterization. To accurately extract the intrinsic transistor s-parameter and noise from the measurement results, various de-embedding techniques have been presented in the last few years [25]-[28]. However, these methods either use a lumped equivalent circuit model or electro-magnetic (EM) simulations to characterize the parasitics. The accuracy of these de-embedding methods at the millimeter-wave band is problematic. Therefore, a general high-frequency de-embedding technique that considers the distributive nature of the parasitics is required. Meanwhile, the impact of substrate coupling on the *ac* and noise behavior should be accurately modeled for circuit analysis and optimization. This also requires such a technique to characterize the substrate.

Another problem is modeling device broadband noise at high frequencies. Discrepancies between modeled and measured noise model parameters have been observed at high frequencies (e.g., R_n, Y_{opt}) [29]. Moreover, the current broadband noise models (e.g., shot noise models) in SiGe HBTs were all developed using the quasi-static assumption. This assumption becomes invalid if the device is operating at ultra-high frequencies, where the changing of noise power density cannot catch up with the changing of base or collector current. Currently, little work has been done to address this problem and thus an accurate noise model for high-frequency application is needed.

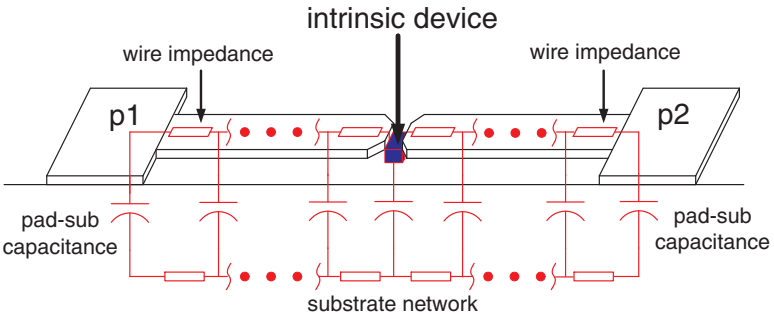

Figure 5: The distributive on-wafer parasitics of a SiGe device.

1.3 Linearity modeling and analysis

Methods of linearity analysis were developed in the 1960s [30]-[32] to simulate the response of nonlinear systems under periodic stimulus (e.g., sinusoidal signals). In integrated-circuit analysis, the three main methods are harmonic balance (frequency domain), periodic steady-state (time domain), and Volterra series (frequency domain). The first two are numerical methods and suited to large-signal simulations. The third is an analytical method and is suitable for small-signal linearity analysis.

In RFIC front-end designs, linearity is another important characteristic [33][34] because it determines the upper limit of the input RF signal power level [22]. One of the figures-of-merit for linearity in a device or an amplifier is the input third-order intercept point (*IIP3*), defined as the intercept point of the extrapolation of the fundamental output and the third-order intermodulation [35]. Figure 6 shows the first-order output power $P_{out,1st}$ and the third-order intermodulation power $P_{out,3rd}$ as functions of input power and the extracted third-order intercept point *IIP3* of a nonlinear amplifier. Observe that the $P_{out,1st}$ increases 20 dB per decade of input power, while the $P_{out,1st}$ increases 60 dB per decade of input power. The third-order intermodulation power should be minimized, since it distorts the output signal. A rule-of-thumb is that $P_{out,3rd}$ should be lower than the noise floor (as shown in Figure 6) so that it is negligible compared to the noise power. For a system with a given gain and *IIP3*, this requirement determines the maximum input power. Figure 6 also indicates that for systems with the same input power and gain, the lower the $P_{out,3rd}$, the higher the *IIP3*, and vice versa (this is shown in Figure 36 in Chapter IV).

A detailed analysis of *IIP3* behavior at both the device level [36] and circuit level [37][38]

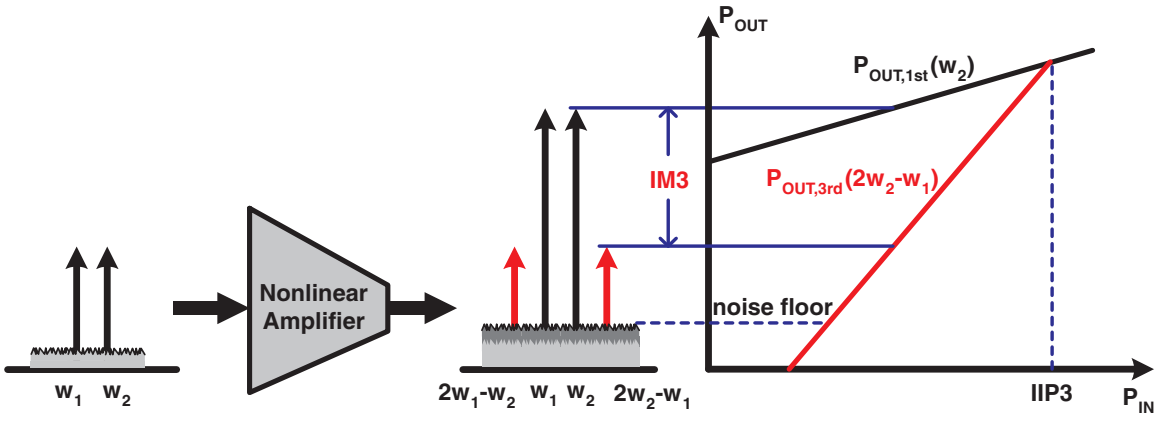

Figure 6: The first-order output power $P_{out,1st}$ and the third-order intermodulation power $P_{out,3rd}$ as functions of input power, and the extracted third-order intercept point *IIP3* of a nonlinear amplifier.

has been presented. However, the analysis of *IIP3* is either too complicated to obtain an explicit form necessary to gain insight into linearity optimization, or has too many simplifying assumptions, leading to a significant loss of accuracy in real applications. In addition, all these analyses assumed that the lumped compact model can accurately capture the nonlinear behavior of the device. However, the model used is hard to verify using current measurement setups (e.g., two-tone, one-tone, or power-sweep measurements). Therefore, developing a simple, yet accurate, and easy to verify model of linearity is another motivation for this dissertation.

1.4 EDA tools for RF circuits

Since researchers at the University of California at Berkeley invented PSPICE in the early 1970s, various integrated circuit simulation tools have been developed to catch up with the growth of process technology (e.g., HPSICE, MNS, Spectre). Nowadays, the most widely used electrical-design-automation (EDA) tools in both industry and at universities are Cadence and HP-ADS. However, these tools suffer several limitations.

First, the nonlinear simulation engines used in these two simulators are periodic steady-state (Cadence) and harmonic balance (HP-ADS) methods. Because of the associated computer rounding errors, both methods lose accuracy as the RF signal power level becomes extremely small compared with the *dc* bias level. Also, both methods are subject to convergence problems at a very high RF signal power level. Therefore, both methods are unsuited to *IIP3* simulations because the user has

to carefully select the RF signal power range to obtain robust and accurate results.

Second, there are limitations on device noise modeling in both tools. To support user-defined device modeling, Cadence uses the Verilog-A analysis tool and HP-ADS uses the symbolic-defined device (SDD) tool. The user cannot define the correlations between different noise sources using the provided tools. These correlations, however, are important when modeling noise behavior at high frequencies and should not be neglected.

Third, there are limitations in the circuit optimization methods available to RF designers. The Cadence tool does not support circuit optimization. There are optimization tools in HP-ADS. Using these tools, the user can specify optimization target(s), strategy, and the design variables' ranges and then optimize the design. However, a complex RFIC design (e.g., LNA) often requires a number of optimization targets (e.g., noise figure *NF*, *IIP3*, gain, input match S_{11}, and output match S_{22}). In this case, the user cannot obtain an optimum design by simply specifying all the targets and including in the whole design space because the workstation cannot afford such enormous numerical calculations. Instead, the user can optimize only one or two targets in a design subspace and then increase the set of optimization targets step by sstep in a larger design subspace. This optimization process is both tedious and time consuming.

Moreover, those tools offer few options in device-level analysis of the linearity and noise performance. However, as device-process technology improves, device engineering of SiGe HBTs is becoming a key solution in RFIC design. In EDA tools, the capability of device-level analysis and optimization in EDA tools becomes more and more necessary.

Therefore, an EDA tool that can accurately calculate and model the linearity and noise and can intelligently optimize RF circuits at both the device and circuit level will be a significant asset for modern RFIC design.

1.5 Conclusion

From a device standpoint, the rapid growth of SiGe HBTs technology offers a cheap and robust solution in RFIC. However, this state-of-the-art device technology has brought forth new problems: characterizing and modeling novel device physics effects (such as the Ge-grading effect, the high-J_C barrier effect, etc.) and their consequences for noise and linearity performance.

From a circuit analysis and design standpoint, device-level analysis is required to fully exploit the potential of device engineering of SiGe HBTs in IC design. This raises a new demand in EDA tool development.

From an *ac* testing standpoint, accurate extraction of the intrinsic device/circuit *ac* and noise parameters is required to support device/circuit analysis. However, this becomes a significant challenge in the millimeter-wave range. Besides the intrinsic device/circuit, the parasitics, such as substrate coupling, should also be modeled and predicted during RF design. An accurate parasitic model is thus required in *ac* characterization.

These issues provide the main objectives of this thesis. Details of the following work will be covered in later chapters:

1. Simulation and modeling of novel device physics effects in SiGe HBTs (Chapter II, also published in [43]).

2. Development of a new EDA tool that decouples the individual physics impacts on noise and linearity behavior in integrated circuits (Chapter III, also published in [36]).

3. Application of the presented EDA tool in RF LNA design and optimization (Chapter IV, also published in [39]).

4. Construction of a new behavioral-model technique to simplify linearity analysis and optimization (Chapter V, also published in [40]).

5. Establishment of a new technique to accurately extract the *ac* and noise parameters of the intrinsic and substrate model in the millimeter-wave range (Chapter VI, also published in [76]).

CHAPTER II

DEVICE SIMULATION AND MODELING

2.1 Introduction to SiGe HBTs

The essential operational differences between SiGe HBTs and Si BJTs are best illustrated by considering a schematic energy band diagram. Figure 7 shows the energy band diagram for a Si BJT and graded-base SiGe HBT [8]. In such a device construction, the Ge content is linearly graded from 0% near the metallurgical emitter-base (EB) junction to some maximum value of Ge content near the metallurgical collector-base (CB) junction. The Ge induced reduction in base bandgap occurs at the EB edge of the quasi-neutral base ($\Delta E_{g,Ge}(x = 0)$), and at the CB edge of the quasi-neutral base ($\Delta E_{g,Ge}(x = W_b)$). This grading of the Ge across the neutral base induces a built-in quasi-drift field ($(\Delta E_{g,Ge}(x = W_b) - \Delta E_{g,Ge}(x = 0))/W_b$) that will impact minority carrier transport.

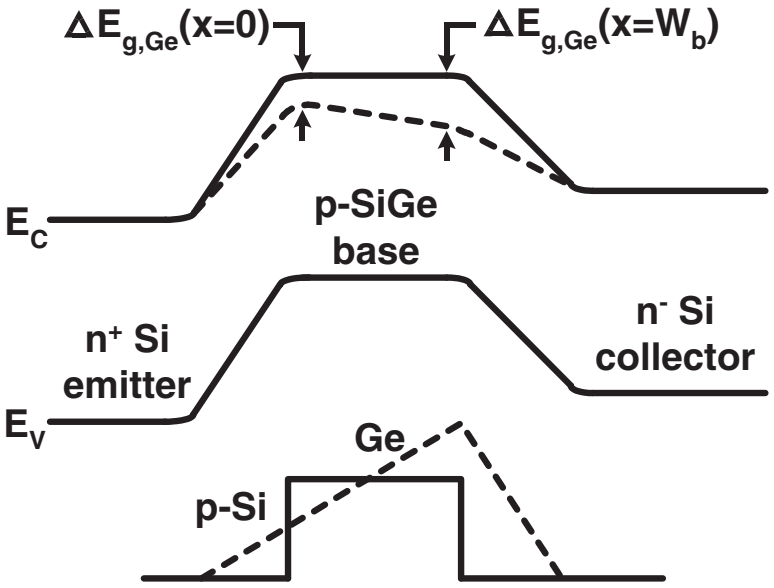

Figure 7: Energy band diagram for a Si BJT and graded-base SiGe HBT in the equilibrium condition.

For simplicity, the ideal graded-base SiGe HBT and the Si BJT with the same, constant doping levels in the emitter, base and collector regions are compared to demonstrate the impact on device

performance. The improvement in current gain β is derived as [8]

$$\frac{\beta_{SiGe}}{\beta_{Si}} \propto \frac{\Delta E_{g,Ge}(grade)/kTe^{\Delta E_{g,Ge}(0)/kT}}{1-e^{-\Delta E_{g,Ge}(grade)/kT}}, \tag{1}$$

where $\Delta E_{g,Ge}(grade) = \Delta E_{g,Ge}(x = W_b) - \Delta E_{g,Ge}(x = 0)$. Note that there is a exponential term $e^{\Delta E_{g,Ge}(0)/kT}$ in the expression, indicating that a huge enhancement in the current gain can be obtained by increasing a few kTs in $\Delta E_{g,Ge}(0)$.

The Ge bandgap narrowing induced built-in field also decreases the base transit time τ_b, which dominates the total transit time in Si BJTs. The ratio of the τ_b between the SiGe HBT and the Si BJT is written as [8]

$$\frac{\tau_{b,SiGe}}{\tau_{b,Si}} \propto \frac{kT}{\Delta E_{g,Ge}(grade)} \cdot \left\{ 1 - \frac{kT}{\Delta E_{g,Ge}(grade)} \left[1 - e^{-\Delta E_{g,Ge}(grade)/kT} \right] \right\}. \tag{2}$$

When $\Delta E_{g,Ge}(grade)$ increases, the τ_b decreases. Therefore, the peak cut-off frequency $f_{T,max}$ of the HBT increases.

As demonstrated above, the current gain and speed of SiGe HBTs significantly increase simply by applying the bandgap-engineering technique. This approach widens the device design range so that other performance factors (e.g., noise, early voltage) can also be improved. For example, the maximum base doping density is traditionally limited by the β requirement in the BJTs–the higher the base doping, the lower the β. Using the SiGe HBTs technology, one can raise the base doping density by one or two levels of magnitude without β degradation. Consequently, the base resistance and minimum noise figure (see Chapters III and IV) of the SiGe HBTs can be much lower than that of the Si BJTs.

Moreover, these ratios are strongly T-dependent: at lower temperatures, the difference between SiGe HBTs and Si BJTs are more obvious. In fact, for SiGe HBTs, the cryogenic performance is even better than the room-temperature performance, opening the way to a variety of niche circuit applications involving cryogenic electronics (i.e., those operating at 77 K or lower) [41].

There are, however, some unique non-ideal effects in SiGe HBTs. The Ge grading effect [42], which arises as a result of the graded Ge profile across the EB space charge region, degrades the β flatness in the medium-bias range. And the high-injection heterojunction barrier effect [43], which is due to the Kirk-effect-induced [44] exposure of the SiGe/Si heterojunction into the neutral-base region, severely degrades both *dc* and *ac* device performance in the high-injection-bias range.

Deeper physical insights into the pros and cons of SiGe HBTs are required for both device-technology and integrated-circuit applications. In this chapter, we investigate the underlying mechanism of the high-injection heterojunction barrier effect and its impact on the transit time of SiGe HBTs, using 2-D device MEDICI simulations. Then, we will derive a physics-based analytical model to capture this effect for circuit analysis.

2.2 2-D simulation tools

To quantitively study the impact of device physics and process technology on state-of-the-art device characteristics, one needs to use Technology-Computer-Aided-Design (TCAD) tools. 2-D MEDICI simulations are used in this thesis.

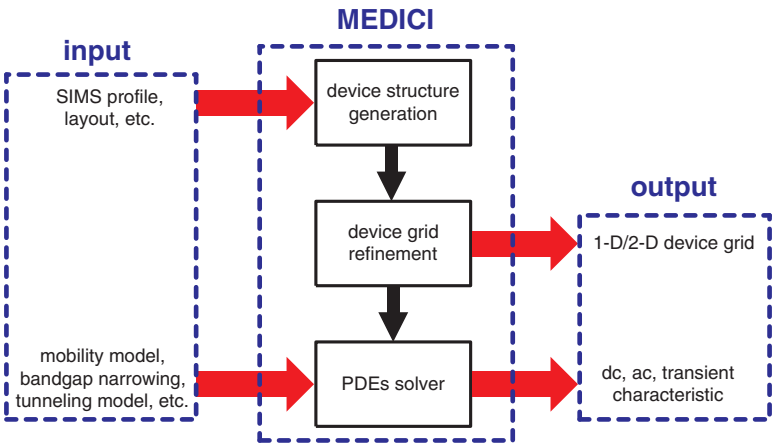

Figure 8: MEDICI simulation procedure.

Figure 8 illustrates the device simulation procedure using MEDICI, which is:

1. Build a device (1-D/2-D) similar to the fabricated one using designed and measured process data (e.g., layout, SIMS profile);

2. Construct and refine a mesh/grid structure on the device for numerical calculations;

3. Specify and solve the partial differential equations (PDEs) using appropriately calibrated physics parameter models.

The partial differential equations include: Poisson's equation (3) and equations for the electron/hole current continuity (4)(5), lattice heat continuity (6), and electron/hole energy balance (7)(8),

as listed below. [45]

$$\varepsilon\nabla^2\psi = -q(p - n + N_D^+ - N_A^-) - \rho_S \quad (3)$$

$$\frac{\partial n}{\partial t} = \frac{1}{q}\vec{\nabla}\cdot\vec{J_n} - U_n = F_n(\psi, n, p), \quad (4)$$

$$\frac{\partial p}{\partial t} = \frac{1}{q}\vec{\nabla}\cdot\vec{J_p} - U_p = F_p(\psi, n, p), \quad (5)$$

$$\rho c\frac{\partial T}{\partial t} = H + \vec{\nabla}\cdot\left(\lambda(T)\vec{\nabla}T\right), \quad (6)$$

$$\vec{\nabla}\cdot\vec{S_n} = \frac{1}{q}\vec{J_n}\cdot\vec{E} - \frac{3}{2}\left[n\frac{u_n - u_0}{\tau_{wn}} + \frac{\partial(nu_n)}{\partial t}\right] - \frac{1}{q}E_g G_n^{II} + H_n^R, \quad (7)$$

$$\vec{\nabla}\cdot\vec{S_p} = \frac{1}{q}\vec{J_p}\cdot\vec{E} - \frac{3}{2}\left[p\frac{u_p - u_0}{\tau_{wp}} + \frac{\partial(pu_p)}{\partial t}\right] - \frac{1}{q}E_g G_p^{II} + H_p^R, \quad (8)$$

where ψ, n, p, T, u_n, and u_p are the potential, electron/hole density, temperature, and electron/hole energy, respectively. These are the six unknown distribution functions to be solved for. The other variables (except for the physics/material coefficients such as q, ρ, c, etc.) in the equations are determined by the six functions. The relations among these variables and the six functions are expressed using physics parameter models such as carrier-mobility models, heavy-doping bandgap-narrowing models, carrier-lifetime models, impact-ionization models, tunneling models, etc.

To solve the PDEs on a computer, they must be discretized using the generated device grid. The continuous functions of the PDEs are represented by vectors of function values at the nodes, and the differential operators are replaced by suitable difference operators. Figure 9 shows an example of the discretization of a continuous differential operator to a difference operator in a 1-D device structure. Using the discretization technique, MEDICI solves for $6N$ real values instead of solving the six functions, where N is the number of grid points.

The physics parameter models mentioned above are calibrated using the reported measurement data. Moreover, they can be turned on/off or changed (e.g., by using different models or model parameters) during the simulation. Hence, various device physics effects can be decoupled and the consequences on device behavior can be predicted.

In the same manner as for a circuit simulator (PSPICE, HP-ADS, Spectre), MEDICI offers *dc*, *ac*, and transient analysis. One can mimic the *dc*, *ac*, and transient measurements in simulations and compare the results with those obtained experimentally. Using MEDICI, one can modify the process (such as the doping profile, the Ge content, the device geometry, etc.) and gain the insights

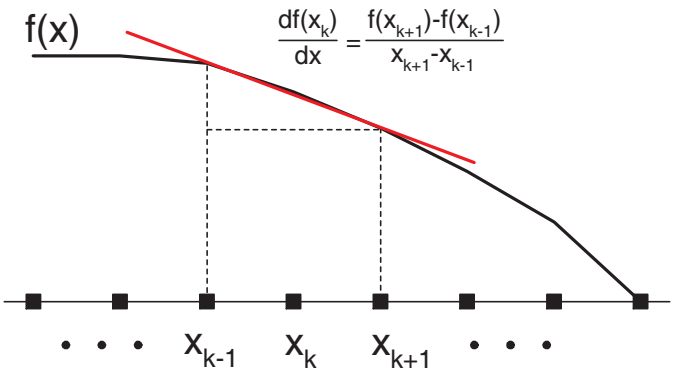

Figure 9: An example of discretization of a continuous differential operator to a difference operator in a 1-D device structure.

into the impact of process technology on device performance. Therefore, MEDICI is a good tool for connecting the device technology, physics, and characteristics. Then we apply this tool to the analysis of the high-injection effects in SiGe HBTs.

2.3 Impact of high-J_C HBE on transit time

In SiGe HBTs, there are two high-injection effects that significantly degrade the f_T and transit time: the conventional Kirk effect and the heterojunction barrier effect. The main contribution is due to the increases in neutral-base charge, since this charge dominates the total transit time at high injection. While semi-empirical or empirical models of the neutral-base charge in SiGe HBTs have been presented [46][47], these models do not accurately fit the measured data at very high J_C for SiGe HBTs, especially for high-breakdown voltage devices.

Figure 10 shows the measured and fitted cut-off frequency as a function of J_C. Observe that the f_T data decreases more rapidly than the HICUM model suggests at J_C above 2.0 $mA/\mu m^2$. The VBIC model also fails to accurately capture the behavior. This model failure also causes discrepancies in the *ac* device parameters used in RF circuit design (e.g., C_{BE}). Figure 11 shows C_{BE} as a function of J_C derived using HICUM compared with measured data. The model clearly underestimates C_{BE} at J_C above 2.0 $mA/\mu m^2$.

Those deviations may cause problems in large-signal applications, such as power amplifiers. Figure 12 shows the output characteristics with $R_{load} = 280\ \Omega$. Even though the *dc* bias of the device is in the low injection range, the device will reach high injection operation during a transient

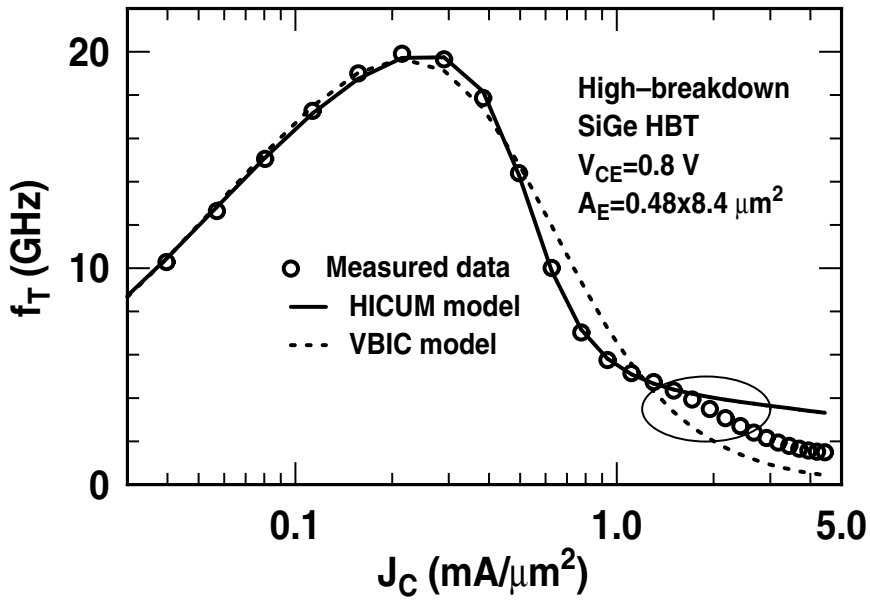

Figure 10: Comparison of measured $f_T - J_C$ characteristics with the HICUM and VBIC compact models.

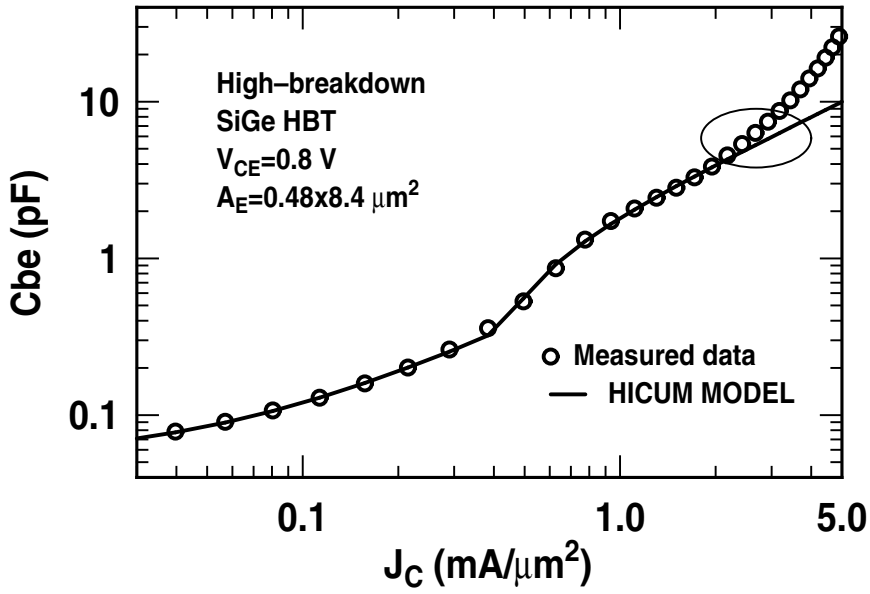

Figure 11: Comparison of measured $C_{BE} - J_C$ characteristics with the HICUM model.

sweep ($I_C = 8\ mA$ and $V_{CE} = 0.8\ V$ in the ellipse). One thus needs to consider the C_{BE} and changes in stored charge (f_T) in this operational range.

Further study shows that this model-to-data discrepancy is caused by the heterojunction barrier effect. In SiGe HBTs, the transition from a narrow gap SiGe base layer to the larger bandgap Si

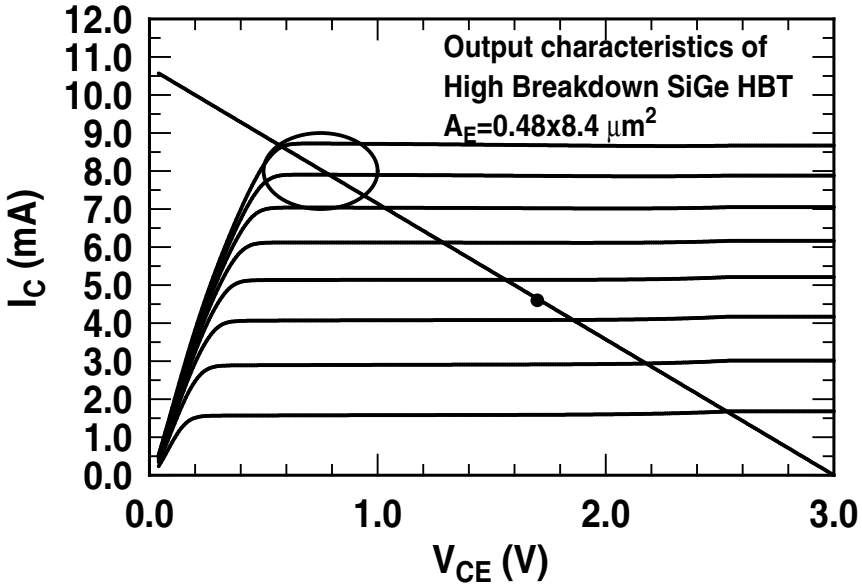

Figure 12: Output characteristics of a high-breakdown SiGe HBT.

collector layer introduces a valence band offset at the hetero-interface. Since this barrier is masked by the band bending in the CB depletion region during low-injection operation, it has a negligible effect on the device characteristics. At high injection, however, the collapse of the original CB electric field at the hetero-interface reveals the barrier, which opposes the hole injection into the collector [48][49]. The hole pile-up that occurs at the hetero-interface induces a conduction band barrier that opposes the flow of electrons into the collector. This causes an increase in the stored base charge, resulting in the sudden decrease of both f_T and f_{max}. Since the amount of holes that are blocked by the barrier is a thermally-activated function of the barrier height, it is expected that this barrier effect will show a much more pronounced effect on the performance of the SiGe HBTs at reduced temperatures [51].

In most compact models, the Kirk effect and barrier effect are lumped into one function, assuming the Kirk effect and barrier effect occur simultaneously. This assumption, however, is no longer valid when the SiGe/Si heterojunction is located either in the neutral base region or deeper in the epitaxial collector. The latter case, for instance, might be found in SiGe HBTs optimized for high breakdown voltage [50]. Figure 13 shows the pushed-out base caused by the Kirk effect at different J_C values in a SiGe HBT with a deep heterojunction. The upper part of the diagram is the Ge profile. One finds that the neutral base is not subjected to the SiGe/Si heterojunction when the Kirk

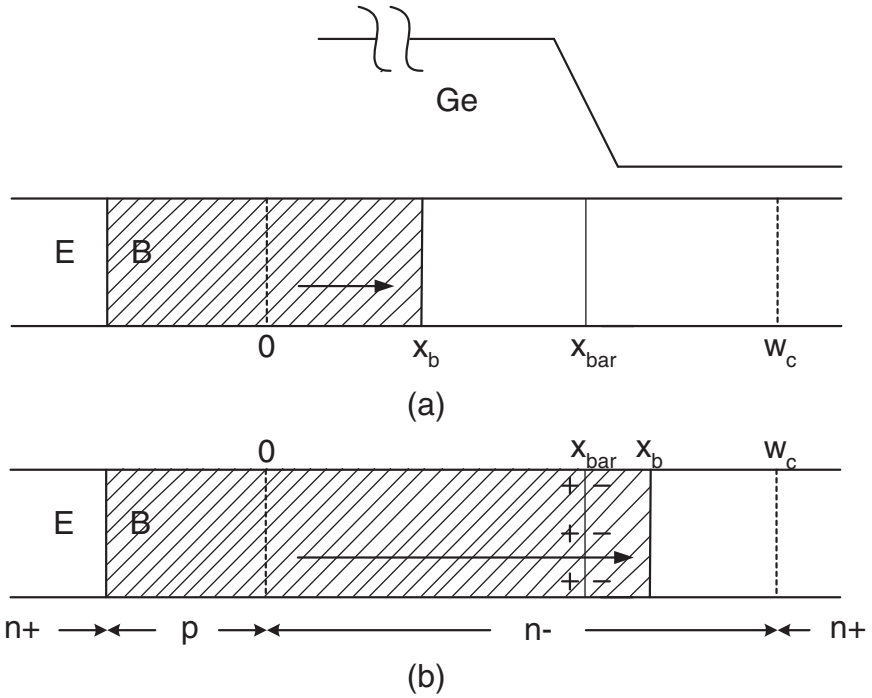

Figure 13: Illustration of pushed-out base caused by Kirk effect at different J_C in a SiGe HBT: (a) pushed-out base does not cover the heterojunction; (b) pushed-out base covers the heterojunction.

effect onset occurs. That is, the barrier effect is significant only when J_C is much higher than the critical onset current density of the Kirk effect.

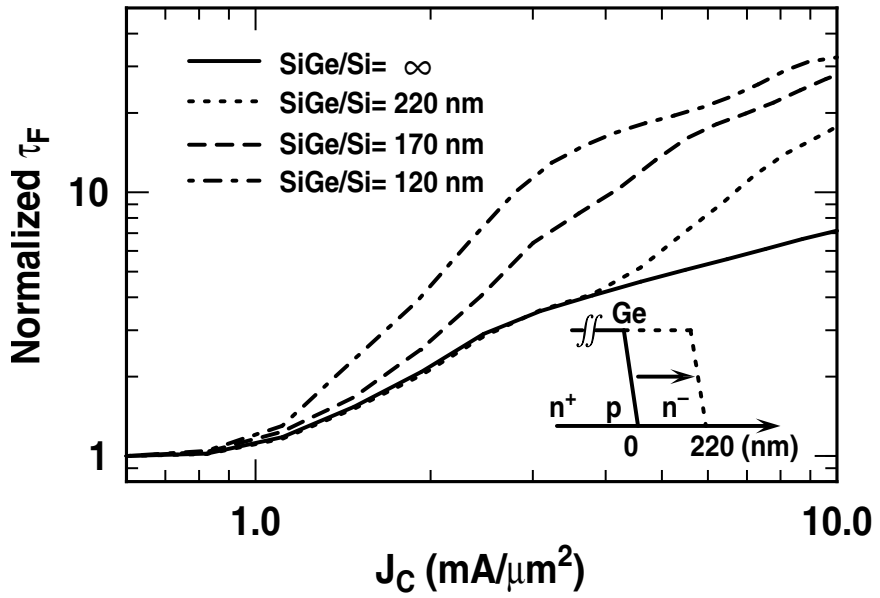

Figure 14: Transit time as a function of J_C for different barrier locations, calculated using MEDICI.

Figure 14 shows simulated values of τ_F as a function of J_C for different heterojunction locations, where $x = 0$ implies the SiGe/Si transition is at the metallurgical CB junction. For a shallow heterojunction location (i.e., 120 nm), the barrier effect is significant when the Kirk effect occurs. However, for a deeper barrier location (i.e., 220 nm), the barrier effect is not significant, since the pushed-out base caused by the Kirk effect has not yet reached the SiGe/Si heterojunction. To accurately capture this phenomenon, a new transit time model that decouples the two effects is thus needed.

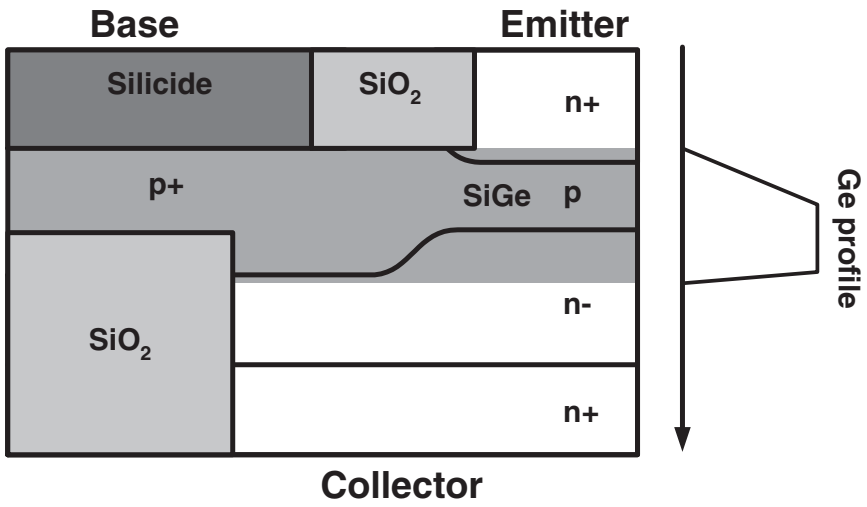

Figure 15: Cross-section of the SiGe HBT used in the MEDICI simulations.

Calibrated 2-D MEDICI simulations is used to obtain deeper insights into how the neutral-base charge density depends on J_C. Figure 15 shows a cross-section of the SiGe HBT device used in the simulations. The structure is identical to current state-of-the-art technology. The doping profile was taken from measured SIMS data.

To better understand the impact of the SiGe barrier effect on the device characteristics, we simulated three types of transistors: a Si BJT, a SiGe HBT with an infinitely deep heterojunction, (i.e., extended into the sub-collector), and a SiGe HBT with a 220 nm deep heterojunction (as measured from the metallurgical CB junction). The trapezoidal Ge profile has a peak Ge content of 10%, and is graded linearly across the neutral base, but is constant in the CB junction, as depicted in the upper part of Figure 13.

Figures 16 and 17 show the simulated electric field and energy band diagrams as functions of

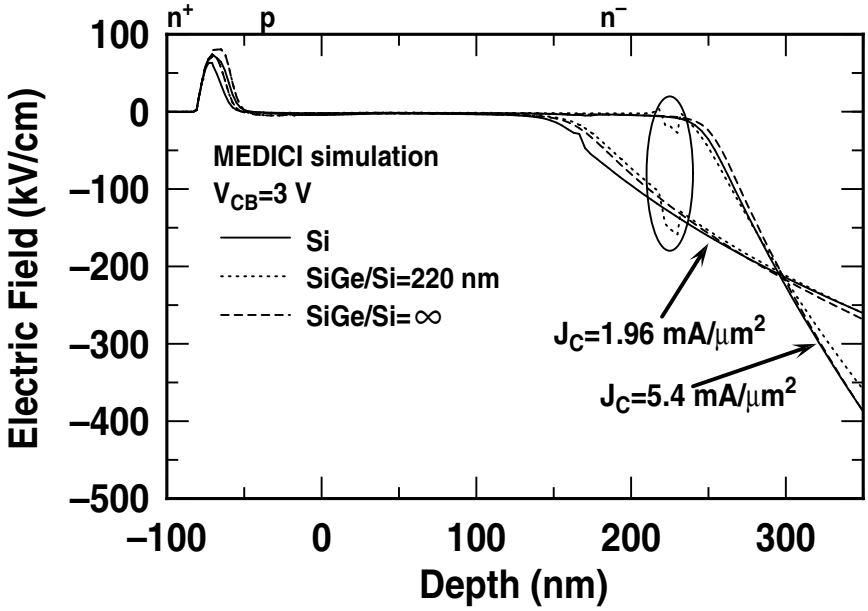

Figure 16: Simulated electric field as a function of depth for a Si BJT and two SiGe HBTs, with J_C values at which the barrier effect both is and is not significant. The metallical BC junction is located at 0 nm.

depth at medium and high J_C for the two SiGe HBTs. A high J_C corresponds to a large pushed-out base region. In this region, the electric field is low and the band is almost flat. The electron mobility is maximized and is approximately constant across the neutral base region (Figure 18). Thus, the assumption of charge neutrality in this region is valid.

The electric field and energy band diagram are subjected to the barrier effect: at medium J_C, the SiGe/Si heterojunction is covered by space-charge-region, and there are no holes piled-up to counter the heterojunction-induced electric field. Thus, there is a large enhanced electric field and no barrier induced in E_C. At high J_C, however, the SiGe/Si heterojunction is exposed in the neutral base. Here, the heterojunction-induced electric field is suppressed by the piled-up holes, and a barrier in the E_C results.

Considering a 1-D transistor, the transit time can be modeled as

$$\tau_F = \frac{q\Delta N}{\Delta I_C} \approx \frac{q\int \Delta n(x)dx}{\Delta J_C} = \frac{q\int_E \Delta n(x)dx}{\Delta J_C} + \frac{q\int_B \Delta n(x)dx}{\Delta J_C} + \frac{q\int_C \Delta n(x)dx}{\Delta J_C} = \tau_e + \tau_b + \tau_c \quad (9)$$

Hence, $\Delta n/\Delta J_C$ as a function of depth determines the dependence of τ_F on J_C.

Figures 19 and 20 show $\Delta n/\Delta J_C$ as a function of depth for the three transistors when the barrier effect is, and is not significant, respectively. The τ_e, τ_b, and τ_c can be obtained by the integration

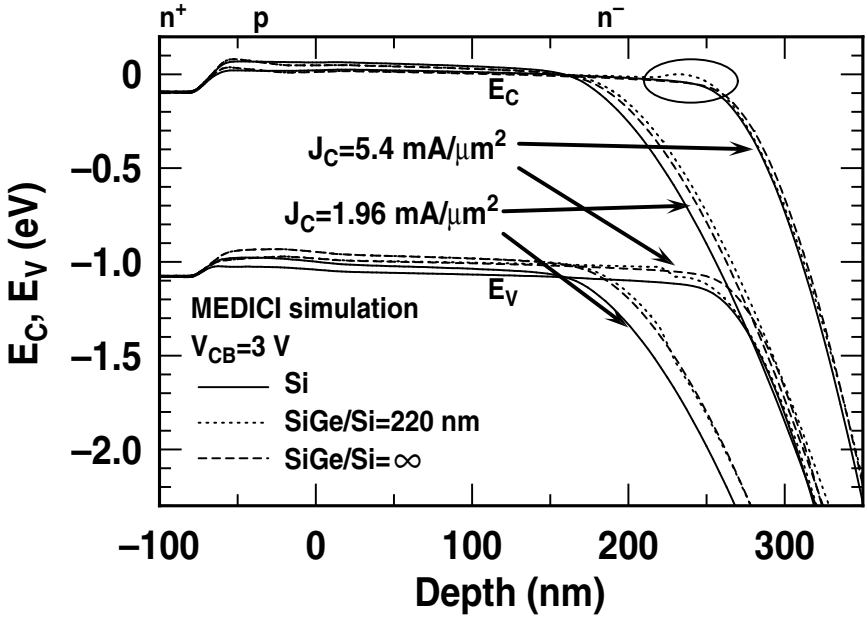

Figure 17: Simulated E_C and E_V as functions of depth for a Si BJT and two SiGe HBTs, with J_C values at which the barrier effect both is and is not significant.

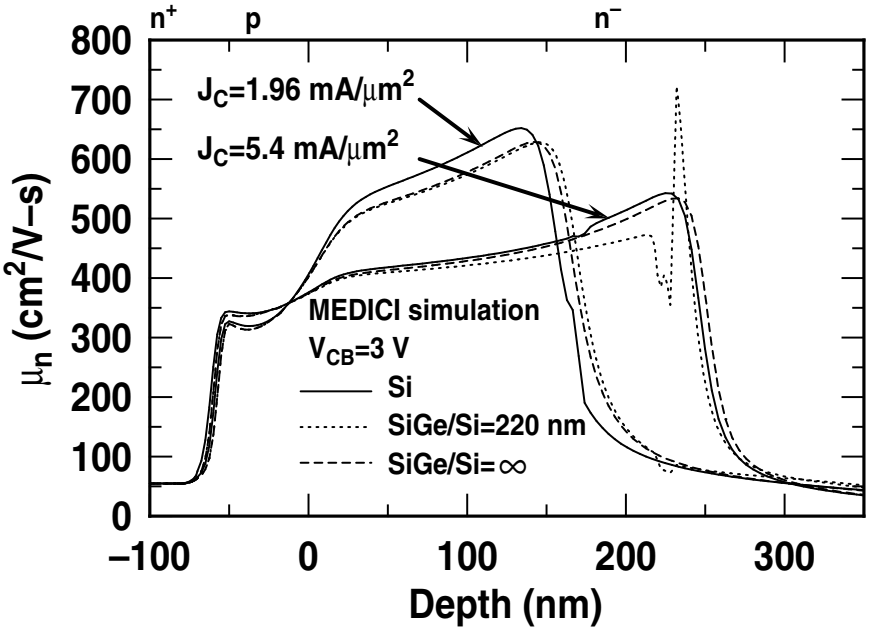

Figure 18: Simulated electron mobility as a function of depth for a Si BJT and two SiGe HBTs, with J_C values at which the barrier effect both is and is not significant.

shown in (9). Since the barrier effect occurs when the heterojunction is located in the pushed-out base, it only changes τ_b. As shown in Figures 19 and 20, when the SiGe HBT is not subjected

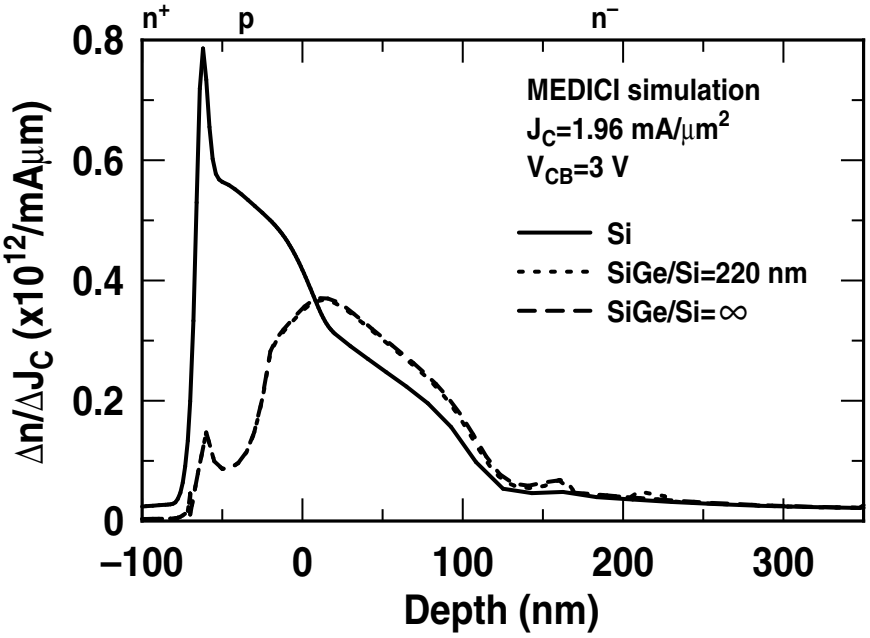

Figure 19: Simulated $\Delta n/\Delta J_C$ as a function of depth for a Si BJT and two SiGe HBTs, with a J_C value at which barrier effects are negligible.

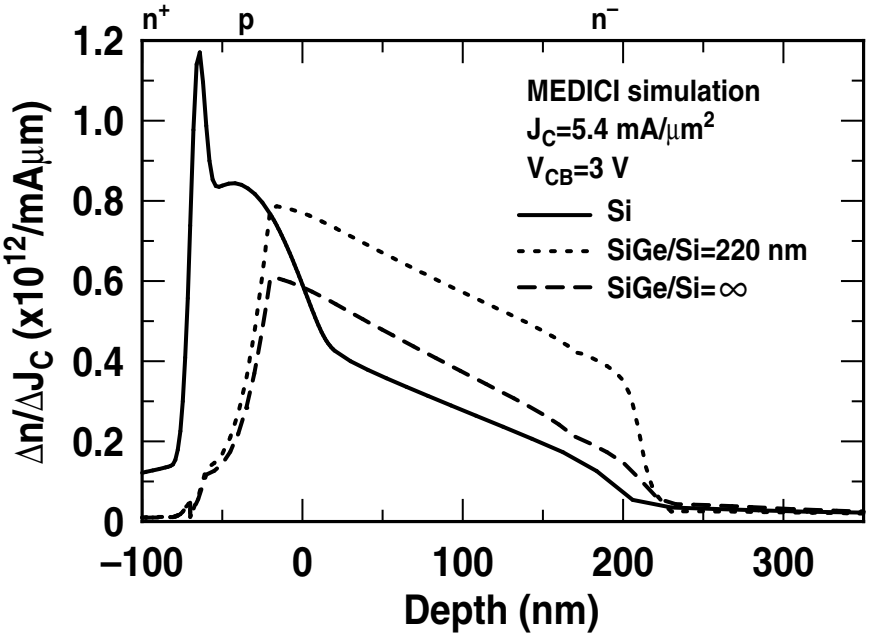

Figure 20: Simulated $\Delta n/\Delta J_C$ as a function of depth for a Si BJT and two SiGe HBTs, with a J_C value at which barrier effects are significant.

to the barrier effect, the shape of $\Delta n/\Delta J_C$ in the neutral base is triangular. However, when the SiGe HBT is subjected to the barrier effect, the shape of $\Delta n/\Delta J_C$ in the neutral base is trapezoidal.

Moreover, the slope of the trapezoid is approximately equal to the slope of the triangle under the same bias conditions. A model of $\Delta\tau_b$ caused by the barrier effect can be derived based on these characteristics, as shown below.

2.4 Model implementation

Based on the simulations, the barrier effect only changes τ_b. Thus, the new transit time model can be written as $\tau_{f,new} = \tau_{f,nobar} + \Delta\tau_b$, where $\Delta\tau_b$ is the increase in transit time due to the barrier effect. An existing compact model (HICUM in this case) can be applied to model $\tau_{f,nobar}$.

Assuming that $\beta >> 1$ and J_p is negligible, the relationship between the charge in the neutral base and the electron flow J_n can be written as [52]

$$-J_n \approx \mu_n n \frac{d(E_{Fn} - E_{Fp})}{dx} = \mu_n n \frac{d\Delta E_F}{dx} \tag{10}$$

where E_{Fn} and E_{Fp} are quasi-Fermi levels for the electrons and holes, respectively, μ_n is the mobility of the electrons, and x is the depth. At high J_C, when the Kirk effect occurs, the density of electrons n near the pushed-out base is much higher than the local collector doping density N_{dc}, suggesting that in this region $n \simeq (n - N_{dc}) \simeq p$. Figure 21 shows the electron and hole distribution at different J_C using MEDICI simulations, confirming that indeed $n \simeq p$ in the pushed-out base.

Thus, n as a function of depth in the pushed-out base can be derived under high injection conditions as (assuming that $n >> N_{dc}$)

$$n(n - N_{dc}(x)) \simeq n_{ib}^2(x) e^{\Delta E_F/kT} = n_i^2 e^{\Delta E_{g0}/kT} e^{(\Delta E_F + x\Delta G_{Eg,Ge})/kT} \tag{11}$$

or

$$n \simeq n_i e^{\Delta E_{g0}/2kT} e^{(\Delta E_F + x\Delta G_{Eg,Ge})/2kT} \tag{12}$$

where n_i is the intrinsic carrier concentration for zero bandgap narrowing, and $\Delta G_{Eg,Ge} = d(\Delta E_{g,Ge})/dx$ is the gradient of the bandgap narrowing caused by the Ge profile (i.e., $\Delta G_{Eg,Ge}$ is constant assuming that the bandgap changes as a linear function of depth). Figure 22 shows the piecewise-linear model for the Ge profile and bandgap narrowing as functions of depth in a deep heterojunction Ge profile, $\Delta G_{Eg,Ge} = 0$ in the epi-collector [4]. Substituting n into (10) and using (12), one obtains

$$\frac{-J_n}{\mu_n} = n\left(\frac{2kT}{n}\frac{dn}{dx} - \Delta G_{Eg,Ge}\right)$$

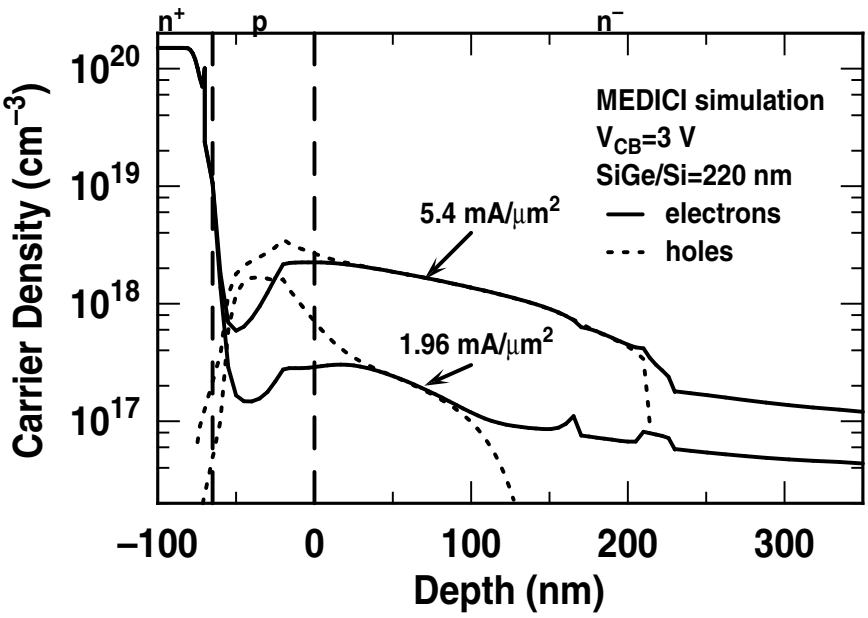

Figure 21: The simulated carrier density as a function of depth at different J_C values in a SiGe HBT.

$$\begin{aligned} \frac{-J_n}{2kT\mu_n} &= \frac{dn}{dx} - \frac{\Delta G_{Eg,Ge}}{2kT} n \\ \frac{-1}{2kT\mu_n} &= \frac{\partial}{\partial x}\left(\frac{\partial n}{\partial J_n}\right) - \frac{\Delta G_{Eg,Ge}}{2kT}\left(\frac{\partial n}{\partial J_n}\right) \end{aligned} \tag{13}$$

In the epitaxial collector region, $\Delta G_{Eg,Ge} = 0$, and thus (13) can be solved as

$$\frac{\partial n}{\partial J_n} = Ax + B \tag{14}$$

where $A = -1/(2kT\mu_n)$, and B is a constant determined by the boundary conditions. Note that at the pushed-out base, μ_n can be viewed as approximately constant (Figure 18). Thus, the distribution of $\Delta n/\Delta J_C$ in the epitaxial collector is a linear function of the depth and the slope is approximately constant.

In general, when $\Delta G_{Eg,Ge}$ is non-zero, (13) can be solved as

$$\frac{\partial n}{\partial J_n} = Ce^{x\Delta G_{Eg,Ge}/2kT} + \frac{1}{\mu_n \Delta G_{Eg,Ge}} \tag{15}$$

where C is a constant determined by the boundary conditions. Since the Ge profiles of most SiGe HBTs with deep heterojunctions are flat in the pushed-out base, we can assume that $\Delta G_{Eg,Ge} = 0$ in the following derivations.

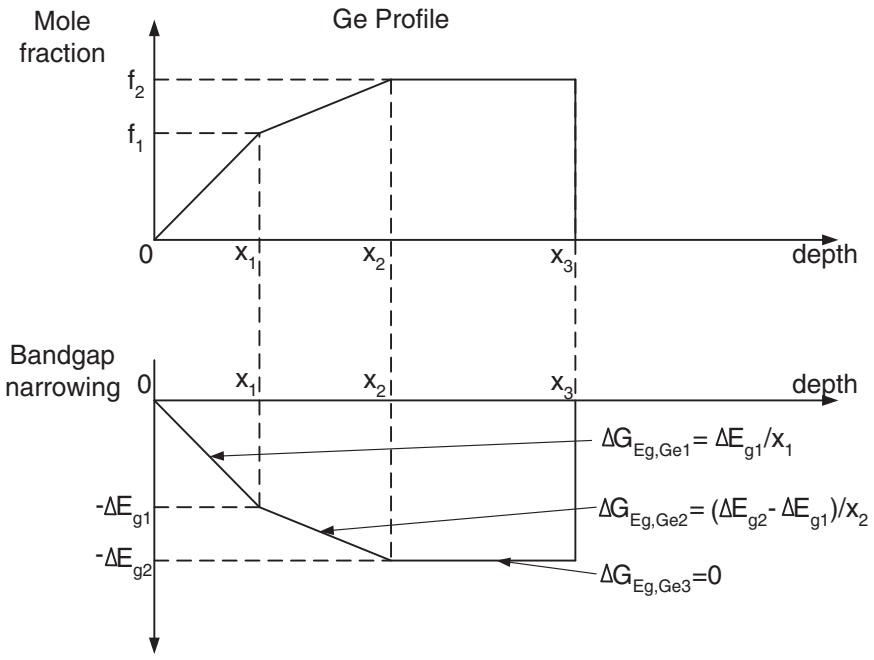

Figure 22: Illustration of piecewise-linear model of Ge mole fraction and ΔE_g as a function of depth.

At the edge of the pushed-out base, J_n is dominated by drift current, and the electric field is so high that the velocity of the electrons is saturated (Figures 16 and 18), and one gets $J_n = qv_{sat}n$. Hence at the edge of the pushed-out base

$$\frac{\partial n(x_b)}{\partial J_n} = \frac{1}{qv_{sat}} \tag{16}$$

where x_b is the boundary at the edge of the pushed-out base.

Near the barrier, the boundary conditions for the electrons are similar to those in the space-charge-region of a pn junction. Hence, one obtains

$$\Delta n(x_{bar}^-) = e^{\Delta E_{bar}/2kT} \times \Delta n(x_{bar}^+) \tag{17}$$

where x_{bar} is the SiGe/Si hetero-interface, and ΔE_{bar} is the barrier height. Here x_{bar}^- is the SiGe/Si hetero-interface on the base side, while x_{bar}^+ is the SiGe/Si hetero-interface on the collector side (Figure 23). Thus

$$\frac{\partial n(x_{bar}^-)}{\partial J_n} = e^{\Delta E_{bar}/2kT} \times \frac{\partial n(x_{bar}^+)}{\partial J_n} \tag{18}$$

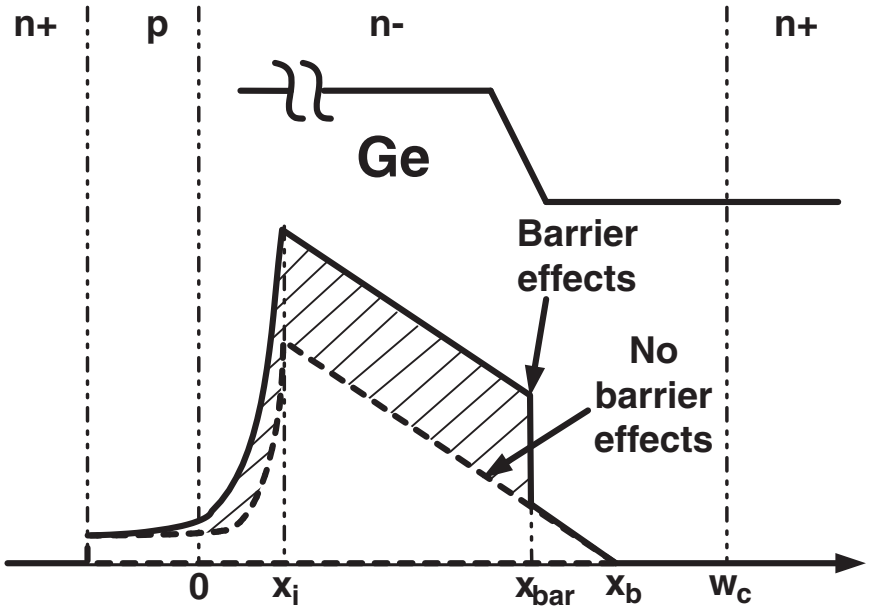

Figure 23: Illustration of $\Delta n/\Delta J_n$ as a function of depth for a SiGe HBT both with and without a barrier effect present. x_i is where the Ge profile begins to flatten, x_{bar} is the location of the heterojunction, and x_b is the edge of the extended base.

Thus, $\Delta n/\Delta J_n$ can be solved using the equations and conditions above. Figure 23 shows $\Delta n/\Delta J_n$ as a function of depth for a SiGe HBT both with and without the barrier effect. The $\Delta\tau_b$ caused by the barrier effect can then be calculated as

$$\Delta\tau_b \approx \begin{cases} 0 & , x_b \leq x_{bar} \\ qx_{bar}|A|(x_b - x_{bar})(e^{\Delta E_{bar}/2kT} - 1) & , x_b > x_{bar} \end{cases} \tag{19}$$

where $A = -1/(2kT\mu_n)$ is the constant defined above. x_b is a function of J_C due to the Kirk effect, and can be written as [47]

$$x_b(J_C) = \begin{cases} 0 & , J_C \leq J_{hc} \\ w_c(1 - J_{hc}/J_C) & , J_C > J_{hc} \end{cases} \tag{20}$$

where w_c is the length of the epi-collector (Figure 23), and J_{hc} the critical onset current for the Kirk effect, which is a function of the epi-collector doping level N_C and V_{CB}. Substituting (20) into (19), one gets

$$\Delta\tau_b(J_C) \approx \begin{cases} 0 & , x_b \leq x_{bar} \\ = \tau_{bb}(1 - J_{bar,c}/J_C) & , x_b > x_{bar} \end{cases} \tag{21}$$

Thus, the deviation of charge ΔQ_b because of the barrier effect is

$$\Delta Q_b(J_C) \approx \begin{cases} 0 & , J_C \le J_{bar,c} \\ \tau_{bb}\left[J_C - J_{bar,c} - J_{bar,c}\ln\left(\frac{J_C}{J_{bar,c}}\right)\right] & , J_C > J_{bar,c} \end{cases} \tag{22}$$

where

$$\begin{aligned} \tau_{bb} &= q|A|\left(e^{\Delta E_{bar}/2kT}-1\right)\left(w_c x_{bar} - x_{bar}^2\right) \\ &= \frac{w_c^2}{2D_n}\left(e^{\Delta E_{bar}/2kT}-1\right)\left[\frac{x_{bar}}{w_c} - \left(\frac{x_{bar}}{w_c}\right)^2\right] \end{aligned} \tag{23}$$

$$J_{bar,c} = \frac{J_{hc}}{(1 - x_{bar}/w_c)} \tag{24}$$

Two additional parameters (x_{bar} and ΔE_{bar}) are used to model $\Delta\tau_b$ caused by barrier effect. τ_{bb} and $J_{bar,c}$ are two useful model fitting parameters, where τ_{bb} is the $\Delta\tau_b$ at $J_C = +\infty$, and $J_{bar,c}$ is the critical onset current density for the barrier effect. One can also define a smooth function to approximate (22) [47]

$$\Delta\tau_b = \tau_{bb}\frac{i_r + \sqrt{i_r^2 + a}}{1 + \sqrt{1 + a^2}} \tag{25}$$

where $i_r = 1 - J_{bar,c}/J_C$, and a is a empirical constant applied to smooth the transition region.

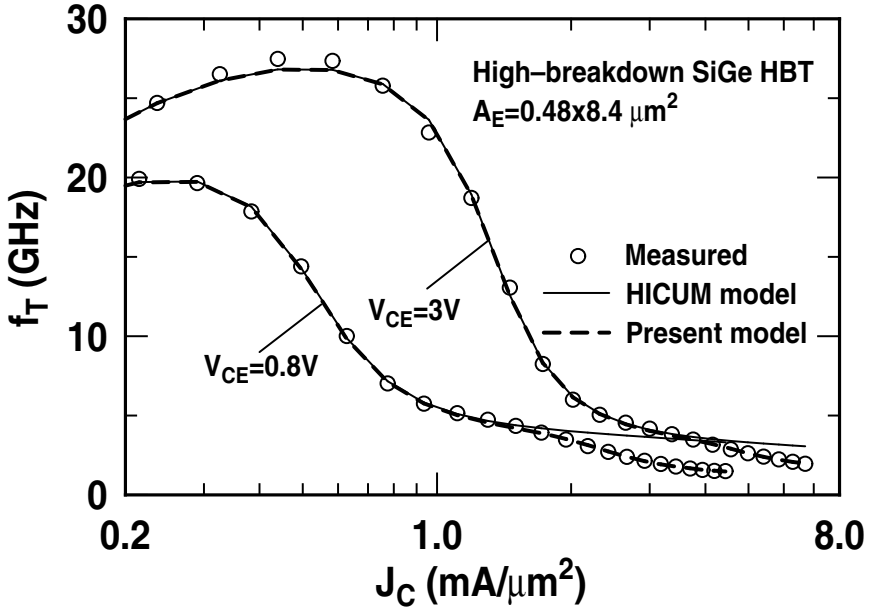

Figure 24: The measured and modeled cut-off frequency as a function of J_C using both the present model, and the default HICUM model.

This new model of the change in $\Delta\tau_b$ caused by the barrier effect can be added to existing compact models such as HICUM by introducing two additional parameters: x_{bar} and ΔE_{bar}. Figure 24 compares the cut-off frequency as a function of J_C using the present transit time model implemented in HICUM, and the original HICUM result versus the measured data. Table 1 shows the key fitting parameters used in the present model. The first four parameters are identical to the parameters introduced in HICUM[47] and the last two are parameters derived above. For a 1-D model, J_{hc} is identical to $I_{CK}/Area$ in HICUM.

Table 1: Key fitting parameters of the present model

Fitting parameters	$V_{ce} = 0.8\ V$	$V_{ce} = 3\ V$
t_{ef0} (ps)	2.00	2.00
t_{hcs} (ps)	37	37
a_{hc}	0.5	0.5
J_{hc} $(mA/\mu m^2)$	0.8	1.9
t_{bb} (ps)	110	110
$J_{bar,c}$ $(mA/\mu m^2)$	2.3	4.7
a	0.03	0.03

Based on the fitting parameters above, $x_{bar} \approx 260\ nm$ and $\Delta E_{bar} \approx 70\ mV$ can be used for this device model. The comparison shows that the new derived model accurately captures the cut-off frequency and transit time as a function of J_C.

CHAPTER III

SYSTEMATIC ANALYSIS OF NOISE AND LINEARITY

3.1 Compact model and circuit analysis

Accurate modeling of passive/active devices is clearly important to integrated circuit simulation. There are two main approaches to building models: a physics-based analytical method and an experimental-behavior-fitting method. The first approach derives the physical equations of the device using several assumptions or simplifications, and the extracted is referred to as a physics-based model. The second approach uses purely-mathematical functions to fit the experiment data, and the extracted is referred to as a behavioral model.

The physics-based model incorporates the insights of device physics, which is a key to device technology optimization. The impact of the fabrication process on device behavior is thus predictable using this model. However, as the device is applied to extreme conditions such as high frequency or high power, some assumptions or simplifications made in the derivations become invalid. Then, more sophisticated functions and model structure have to be used, thus burdening the circuit analysis. Moreover, as the model becomes more complicated, it is hard to extract and verify the model parameters without using additional testing structures.

Figure 25 shows the equivalent circuit structure of an industry paradigm for physics-based models: the VBIC (vertical bipolar inter-company) model [46]. Each component in the equivalent circuit reflects a physical effect (e.g., R_{CI} models the quasi-saturation effect, I_{gc} models the avalanche multiplication, etc.). By tuning the model parameters of each component in simulation, one can decouple the effect of different device-physics phenomena on the *ac*, *dc*, linearity, and noise performance. There are totally 14 nodes in this model, eleven of which are traditional equivalent-circuit nodes, two are in the excess phase network, and one is in the thermal network.

A behavioral model is easy to extract and verify [53]-[58]. Usually the mathematical expressions adopted in behavior models and model structures are simplified. Hence these models are more convenient for circuit analysis. However, they offer few insights into device physics and are thus

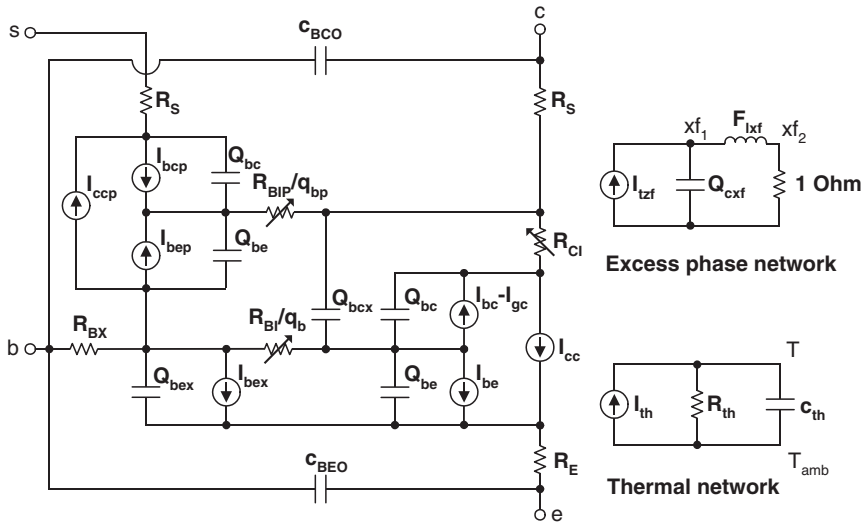

Figure 25: The equivalent circuit of the VBIC model [46].

not favorable in device-level analyses.

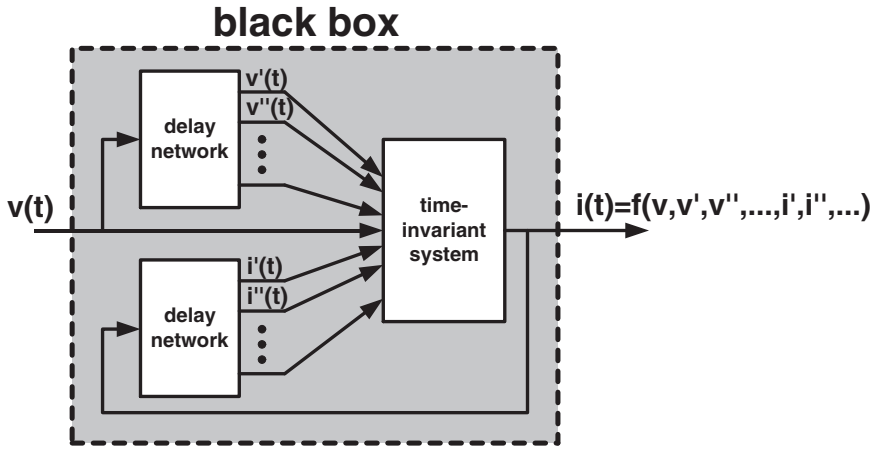

Figure 26: The model structure of a one-port behavioral model [53].

Figure 26 shows the model structure of the one-port behavioral model presented in [53]. In this model, there is actually only one nodal function ($i(t) = f(v(t), v'(t), \cdots, v^{(p)}(t), i'(t), \cdots, i^{(q)}(t))$) which needs to be calculated in simulation. However, it is difficult to link the function with practical device process technology.

Each kind of model has pros and cons, as indicated above. From the calculation algorithm point of view, there is no difference in simulations using either model. Figure 27 shows a general n-node

model in circuit simulation (note if there is only one n-port model in the system, there should also be the n+1-th node as a reference node). The model can be written as

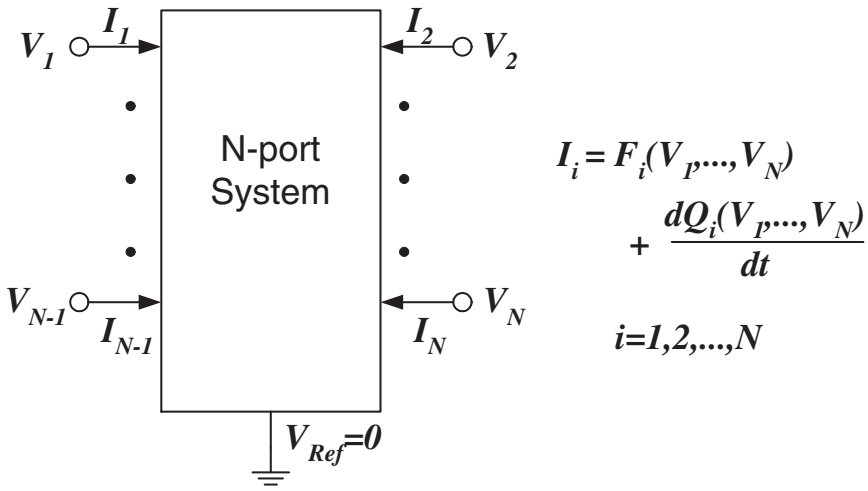

Figure 27: A general n-node model used in circuit analysis.

$$\begin{bmatrix} I_1 \\ I_2 \\ \vdots \\ I_n \end{bmatrix} = \begin{bmatrix} F_1(V_1, V_2, \cdots, V_n) \\ F_2(V_1, V_2, \cdots, V_n) \\ \vdots \\ F_n(V_1, V_2, \cdots, V_n) \end{bmatrix} + \frac{\partial}{\partial t} \begin{bmatrix} Q_1(V_1, V_2, \cdots, V_n) \\ Q_2(V_1, V_2, \cdots, V_n) \\ \vdots \\ Q_n(V_1, V_2, \cdots, V_n) \end{bmatrix}. \tag{26}$$

where $F_1, F_2, \cdots, F_n$ and $Q_1, Q_2, \cdots, Q_n$ are time-invariant nonlinear/linear functions (Later, we also refer to F_i and Q_i as "dc" and "charge" functions, respectively). Without loss of generality, $[I_1, I_2, \cdots, I_n]^T$ and $[V_1, V_2, \cdots, V_n]^T$ are the stimuli and responses of the model instead of the current and voltage as specified in the figure, respectively. The reason for using I and V for the representation is to be compatible with the compact-modified-nodal-analysis (CMNA) [59] approach. In this approach, all sources are converted to current sources as the stimuli, and the nodal voltages are the variables to be solved as the responses.

This general model can be used to represent both a physics model such as the VBIC (which is straight-forward) and the behavioral model discussed above. For the behavioral model, one can

write equation (26) as

$$\begin{bmatrix} i_1 \\ 0 \\ \vdots \\ 0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \begin{bmatrix} f(v_1, v_2, \cdots, v_{p+1}, v_{p+2}, \cdots, v_{p+q+1}) \\ v_2 \\ \vdots \\ v_{p+1} \\ v_{p+2} \\ v_{p+3} \\ \vdots \\ v_{p+q+1} \end{bmatrix} + \frac{\partial}{\partial t} \begin{bmatrix} 0 \\ -v_1 \\ \vdots \\ -v_p \\ -f(v_1, v_2, \cdots, v_{p+q+1}) \\ -v_{p+2} \\ \vdots \\ -v_{p+q} \end{bmatrix}. \quad (27)$$

Therefore, the circuit analysis is generalized to calculate the responses with given stimuli using equation (26). For *dc* analysis, when the system is stable, the time-derivative functions are equal to zero. Therefore, equation (26) becomes

$$\begin{bmatrix} I_1 \\ I_2 \\ \vdots \\ I_n \end{bmatrix} = \begin{bmatrix} F_1(V_1, V_2, \cdots, V_n) \\ F_2(V_1, V_2, \cdots, V_n) \\ \vdots \\ F_n(V_1, V_2, \cdots, V_n) \end{bmatrix}. \quad (28)$$

This can be numerically solved using Newton's method with Gaussian elimination of the Jacobian matrix of $[F_1, F_2, \cdots, F_n]^T$ [45].

For *ac* small-signal analysis, one obtains

$$\begin{bmatrix} i_1 \\ \vdots \\ i_n \end{bmatrix} = \begin{bmatrix} \mathrm{JF}_{11} + s\mathrm{JQ}_{11} & \cdots & \mathrm{JF}_{1n} + s\mathrm{JQ}_{1n} \\ \vdots & \ddots & \vdots \\ \mathrm{JF}_{n1} + s\mathrm{JQ}_{n1} & \cdots & \mathrm{JF}_{nn} + s\mathrm{JQ}_{nn} \end{bmatrix} \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} = \begin{bmatrix} Y_{11}(s) & \cdots & Y_{1n}(s) \\ \vdots & \ddots & \vdots \\ Y_{n1}(s) & \cdots & Y_{nn}(s) \end{bmatrix} \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix}, \quad (29)$$

where $\mathrm{JF}_{ij} = \dfrac{\partial F_i}{V_j}$ and $\mathrm{JQ}_{ij} = \dfrac{\partial Q_i}{V_j}$ are the elements of the Jacobian matrices of $[F_1, F_2, \cdots, F_n]^T$ and $[Q_1, Q_2, \cdots, Q_n]^T$, respectively. $s = j\omega = j2\pi f$ is the operating frequency. This linear equation can be easily solved.

So far, the above derivations show how to implement *dc* or *ac* analysis. In the following sections, the algorithm for systematic noise and linearity calculation will be presented. Only the linear noise analysis and weakly-nonlinear analysis are discussed in this thesis. Unless specified otherwise, the default variables and functions used later are in the frequency-domain rather than the time-domain.

3.2 *Noise calculation algorithm*

There are three types of noise source in the integrated circuit: thermal noise, shot noise, and low-frequency noise. The thermal noise originates from the random (Brownian) motion of the carriers, and the shot noise originates from fluctuations in the carrier diffusion. The physical origin of low-frequency noise is still an open question; one widely held belief is that this noise is related to the trap-induced carrier generation and recombination [60].

A device noise model that characterizes these physics noise sources is required for both device and circuit optimization. Figure 28 shows a general n-port device noise model [61][62]. The device is modeled as a noiseless n-port system with n noise current sources. Unlike conventional signal sources, the noise sources are random signals. Since the broadband noise in a transistor is either Gaussian or Poisson process, which is an ergodic noise process. Therefore, the level of the noise sources and the interactions among different noise sources can be defined using auto-correlation and cross-correlation functions instead of the magnitude and phase. The auto-corvariance and cross-corvariance of the current sources at port i and j can be written as

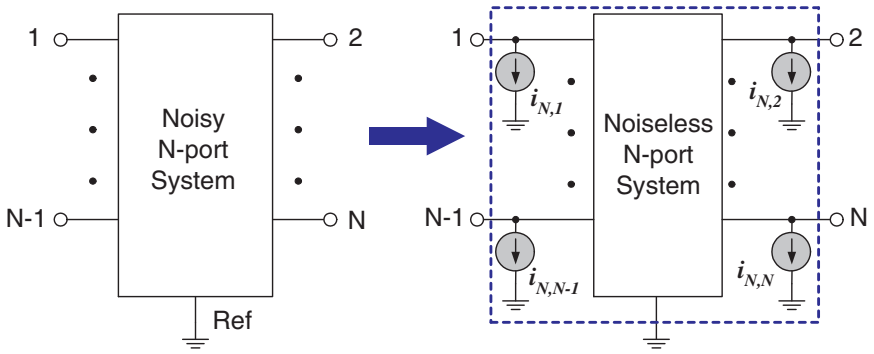

Figure 28: A n-port device noise model used in circuit analysis.

$$\varphi_{ii}(\tau) = \lim_{T\to\infty}\left(\frac{1}{T}\int_{-T/2}^{T/2} i_{n,i}(t)i_{n,i}(t+\tau)dt\right), \tag{30}$$

$$\varphi_{ij}(\tau) = \lim_{T\to\infty}\left(\frac{1}{T}\int_{-T/2}^{T/2} i_{n,i}(t)i_{n,j}(t+\tau)dt\right), \tag{31}$$

where $i_{n,i}(t)$ is the noise-induced random noise fluctuation at port i (which is a measurable time-domain current). φ_{ii} is the time-domain auto-correlation of noise source $i_{n,i}$. φ_{ij} is the time-domain

cross-correlation between $i_{n,i}(t)$ and $i_{n,j}(t)$. The correlation functions characterize the noise behavior as a random process: they indicate how fast the fluctuation decays within a time interval of τ. In simulations, the frequency-domain equivalent noise signal model was used. The frequency-domain noise correlation functions are written as.

$$S_{i_{n,i}i^*_{n,i}}(f) = \int_{-\infty}^{\infty} \varphi_{ii}(\tau) e^{-j2\pi f\tau} d\tau, \tag{32}$$

$$S_{i_{n,i}i^*_{n,j}}(f) = \int_{-\infty}^{\infty} \varphi_{ij}(\tau) e^{-j2\pi f\tau} d\tau, \tag{33}$$

where $S_{i_{n,i}i^*_{n,i}}(f)$ is the current-square spectral density of the auto-correlation of noise source $i_{n,i}$ and $S_{i_{n,i}i^*_{n,j}}(f)$ is the current-square spectral density of the cross-correlation between $i_{n,i}$ and $i_{n,j}$. These are functions of frequency. From the definitions above and using the convolution theory (i.e., appling $t_1 = -t$ and $t_2 - t_1 = \tau$ in equations (30)-(33)), one obtains

$$S_{i_{n,i}i^*_{n,i}}(f) = \int_{-\infty}^{\infty} i_{n,i}(t_1) e^{-j2\pi f t_1} dt_1 \int_{-\infty}^{\infty} i_{n,i}(t_2) e^{j2\pi f t_2} dt_2 = \widetilde{i_{n,i}}(f)\widetilde{i^*_{n,i}}(f), \tag{34}$$

$$S_{i_{n,i}i^*_{n,j}}(f) = \int_{-\infty}^{\infty} i_{n,i}(t_1) e^{-j2\pi f t_1} dt_1 \int_{-\infty}^{\infty} i_{n,j}(t_2) e^{j2\pi f t_2} dt_2 = \widetilde{i_{n,i}}(f)\widetilde{i^*_{n,j}}(f), \tag{35}$$

where $\widetilde{i_{n,i}}(f)$ is the Fourier transformation of the noise source $i_{n,i}(t)$. Observe that the definition of current-square spectral density of the auto-correlation of a noise source is identical to the definition of current-square spectral density of an *ac* signal. Therefore, **the noise source can be treated similarly to an *ac* signal in the frequency-domain**, except that the cross-correlation among different noise sources should also be considered. The average current square of the noise source i in the $(f, f + \Delta f)$ band is written as

$$\overline{i^2_{n,i}} = \int_f^{f+\Delta f} S_{i_{n,i}i^*_{n,i}}(f) df \approx S_{i_{n,i}i^*_{n,i}}(f)\Delta f. \tag{36}$$

For brevity, $S_{i_{n,i}i^*_{n,i}}$ replaces $S_{i_{n,i}i^*_{n,i}}(f)$ when referring to the current-square spectral density. The noise behavior is thus modeled using these current-square spectral densities. In general, the N-port noise sources are equivalent to an $N \times N$ complex matrix:

$$SY_N = \begin{bmatrix} S_{i_{n,1}i^*_{n,1}} & \cdots & S_{i_{n,1}i^*_{n,n}} \\ \vdots & \ddots & \vdots \\ S_{i_{n,n}i^*_{n,1}} & \cdots & S_{i_{n,n}i^*_{n,n}} \end{bmatrix} = \begin{bmatrix} \widetilde{i_{n,1}}\widetilde{i^*_{n,1}} & \cdots & \widetilde{i_{n,1}}\widetilde{i^*_{n,n}} \\ \vdots & \ddots & \vdots \\ \widetilde{i_{n,n}}\widetilde{i^*_{n,1}} & \cdots & \widetilde{i_{n,n}}\widetilde{i^*_{n,n}} \end{bmatrix} \tag{37}$$

In theory, the power spectral densities of the thermal noise and shot noise are not functions of frequency, and they are referred to as broadband noise. The power spectral density of the low frequency noise is a $1/f$-like function of frequency, and it is also referred to as $1/f$ noise. In the RF frequency range, the power spectral density of $1/f$ noise is much smaller than that of the broadband noise. Therefore, the impact of $1/f$ noise is negligible in linear noise analyses at RF frequencies.

To be compatible with the CMNA approach, all noise sources are modeled as current sources here. Figure 29 illustrates a simplified BJT model and the transferred form using noise current sources. Here, $S_{i_b i_b^*}$ and $S_{i_c i_c^*}$ are the current-square spectral density of the base and collector current

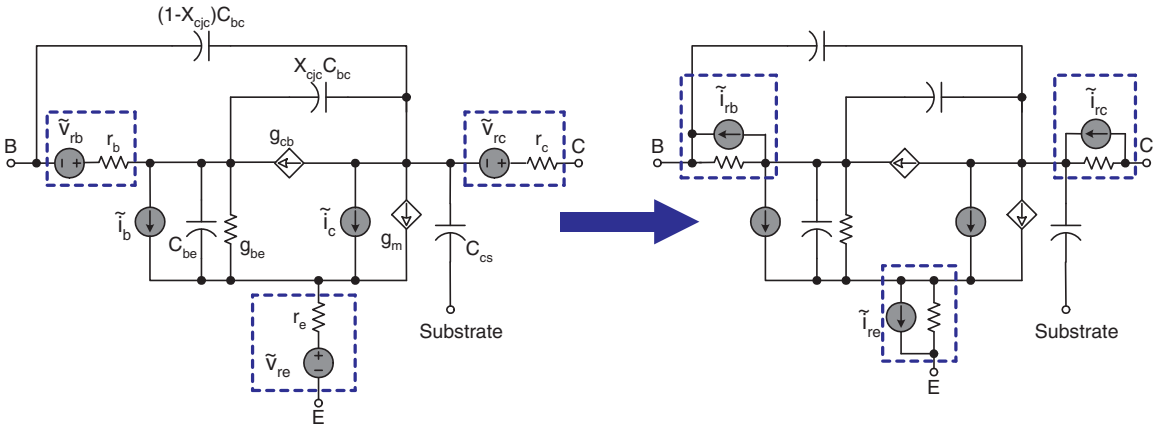

Figure 29: A simplified BJT model and the transferred form using noise current sources.

noise sources, respectively (the shot noise).

$$S_{i_b i_b^*} = 2qI_b, \tag{38}$$

$$S_{i_c i_c^*} = 2qI_c, \tag{39}$$

and $S_{v_{rb} v_{rb}^*}$, $S_{v_{re} v_{re}^*}$ and $S_{v_{rc} v_{rc}^*}$ are the voltage-square spectral density of the thermal noise sources produced by the base, emitter, and collector resistances, respectively. The voltage-square spectral density of the noise voltage source is defined similarly as the current-square spectral density of noise current source, except that the voltage fluctuation instead of the current fluctuation is measured.

$$S_{v_{rx} v_{rx}^*} = 4kTR_x, \tag{40}$$

where "x" refers to "b," "e," or "c" for the base, emitter and collector resistances, respectively. When

transferred into current sources, they are written as

$$S_{i_{rx}i_{rx}^*} = \frac{4kT}{R_x}, \tag{41}$$

These equations are used in conventional compact models, with the assumption that the correlation among each noise source is 0. However, previous research [29] has shown that the correlation between the base and collector noise needs to be considered at high frequencies. This is beyond the capability of the standard circuit analyzers (e.g., HP-ADS, Cadence). Here, we present a systematic noise calculation algorithm that includes the correlation term.

Without loss of generality, the model shown in Figure 28 is adopted for this analysis. In the noise simulation, what we are interested in is the power spectral density of the outside node(s) (namely, the node(s) that can be "forced" or "sensed" in the analysis) for a given circuit structure. The voltage and current of the outside node(s) are denoted v_o and i_o, respectively, and those of the internal nodes are denoted v_i and i_i. Therefore, the Y-matrix described in equation (29) decomposes to

$$Y(s) = \begin{bmatrix} Y_{oo} & Y_{oi} \\ Y_{io} & Y_{ii} \end{bmatrix}. \tag{42}$$

Considering the noise current source, equation (29) is written as

$$\begin{bmatrix} i_o \\ i_i \end{bmatrix} - \begin{bmatrix} \widetilde{i_{n,o}} \\ \widetilde{i_{n,i}} \end{bmatrix} = \begin{bmatrix} Y_{oo} & Y_{oi} \\ Y_{io} & Y_{ii} \end{bmatrix} \cdot \begin{bmatrix} v_o \\ v_i \end{bmatrix}, \tag{43}$$

where $\widetilde{i_{n,o}}$ and $\widetilde{i_{n,i}}$ denote the frequency-domain noise current source at the outside and internal nodes, respectively. The auto-correlation and correlation of $\widetilde{i_{n,o}}$ and $\widetilde{i_{n,i}}$ are

$$SY_n = \begin{bmatrix} SY_{n,oo} & SY_{n,oi} \\ SY_{n,io} & SY_{n,ii} \end{bmatrix}. \tag{44}$$

Then the system can be characterized as the relationship between i_o and v_o, which is written as

$$i_o - \widetilde{i_{n,E}} = i_o - \widetilde{i_{n,o}} + Y_{oi}Y_{ii}^{-1}\widetilde{i_{n,i}} = (Y_{oo} - Y_{oi}Y_{ii}^{-1}Y_{ii})v_o = Y_E(s)v_o, \tag{45}$$

where $Y_E(s) = Y_{oo} - Y_{oi}Y_{ii}^{-1}Y_{ii}$ is the effective Y-matrix representing the $i_o - v_o$ relation. $\widetilde{i_{n,E}} = \widetilde{i_{n,o}} - Y_{oi}Y_{ii}^{-1}\widetilde{i_{n,i}}$ is the effective noise current source. The correlation matrix of $\widetilde{i_{n,E}}$ is calculated as

$$S_{n,E} = \widetilde{i_{n,E}}\widetilde{i_{n,E}^*} = S_{n,oo} - Y_{oi}Y_{ii}^{-1}S_{n,io} - S_{n,oi}(Y_{ii}^*)^{-1}Y_{oi}^* + Y_{oi}Y_{ii}^{-1}S_{n,ii}(Y_{ii}^*)^{-1}Y_{oi}^*. \tag{46}$$

Therefore, the noise characteristics of the outside node(s) can be obtained. For example, the figures-of-merit of the noise in a two-port system are NF_{min}, R_n, and Y_{OPT}. These parameters can be derived from the effective noise correlation matrix SY_2 at the two ports concerned [23]. Using the approach described above, one can directly convert the SY_n to SY_2, and then calculate the noise parameters without neglecting the correlation among the physical noise sources.

3.3 Linearity calculation algorithm

The linearity calculation algorithm adopted here is the Volterra-series approach [32]. Consider the general n-port system illustrated in Figure 27. Using a Taylor expansion and assuming small-signal stimuli, the "dc" functions (F_i) approximate to the polynomials with order less than 4, as shown below

$$F_i = F_{i0} + PF1_i + PF2_i + PF3_i, \tag{47}$$

$$F_{i0} = F_i(V_{10}, V_{20}, \cdots, V_{n0}), \tag{48}$$

$$PF1_i = \sum_{j=1,\cdots,n} \frac{\partial F_i}{\partial V_j} v_j, \tag{49}$$

$$PF2_i = \frac{1}{2} \sum_{j,k=1,\cdots,n} \frac{\partial^2 F_i}{\partial V_j \partial V_k} v_j v_k, \tag{50}$$

$$PF3_i = \frac{1}{6} \sum_{j,k,m=1,\cdots,n} \frac{\partial^3 F_i}{\partial V_j \partial V_k \partial V_m} v_j v_k v_m, \tag{51}$$

where $V_{10}, \cdots, V_{n0}$ are the dc biasing voltages at the n ports. $v_j/v_k/v_m$ is the small voltage variation of node $j/k/m$, respectively. The charge functions (Q_i) are expanded in a similar manner. Then equation (26) is re-written as

$$\begin{bmatrix} I_1 - F_{10} \\ \vdots \\ I_n - F_{n0} \end{bmatrix} = \begin{bmatrix} PF1_1 + \frac{\partial}{\partial t} PQ1_1 \\ \vdots \\ PF1_n + \frac{\partial}{\partial t} PQ1_n \end{bmatrix} + \begin{bmatrix} PF2_1 + \frac{\partial}{\partial t} PQ2_1 \\ \vdots \\ PF2_n + \frac{\partial}{\partial t} PQ2_n \end{bmatrix} + \begin{bmatrix} PF3_1 + \frac{\partial}{\partial t} PQ3_1 \\ \vdots \\ PF3_n + \frac{\partial}{\partial t} PQ3_n \end{bmatrix}, \tag{52}$$

where $I_i - F_{i0}$ is the small net-current variation at the node i. The first term on the right hand side of equation (52) is the linear part of the system, and the second and third terms represent the nonlinear part.

Figure 30 shows the equivalent nonlinear model for an n-port system. The nonlinearities are

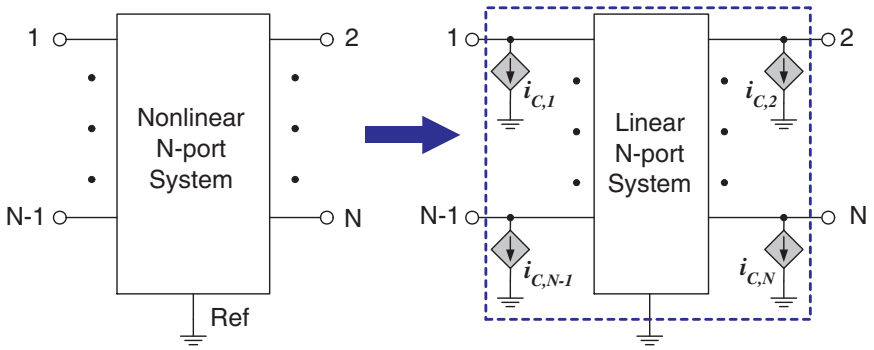

Figure 30: The equivalent nonlinear model for an n-port system.

modeled as voltage-controlled current sources:

$$\begin{bmatrix} i_{C,1} \\ \vdots \\ i_{C,n} \end{bmatrix} = \begin{bmatrix} PF2_1 + \dfrac{\partial}{\partial t} PQ2_1 \\ \vdots \\ PF2_n + \dfrac{\partial}{\partial t} PQ2_n \end{bmatrix} + \begin{bmatrix} PF3_1 + \dfrac{\partial}{\partial t} PQ3_1 \\ \vdots \\ PF3_n + \dfrac{\partial}{\partial t} PQ3_n \end{bmatrix} \tag{53}$$

This approach can be applied to a physics model to decouple the impact of different device physics on linearity. Figure 31 shows a simplified BJT nonlinear model. In this model, six nonlinear sources that are induced by different physical phenomena are used: the nonlinear $I_b - V_{be}$ and $I_c - V_{be}$ relation; the nonlinear EB, CB, and CS capacitance; and the nonlinear avalanche-multiplication current I_{cb}. One can turn on/off each of these nonlinear current sources to determine the dominant contribution.

Returning to the calculation algorithm, for brevity, we denote

$$TF1_{ij} = \mathrm{JF}_{ij} = \frac{\partial F_i}{\partial V_j}, \quad TF2_{ijk} = \frac{\partial^2 F_i}{\partial V_j \partial V_k}, \quad \text{and} \quad TF3_{ijkm} = \frac{\partial^3 F_i}{\partial V_j \partial V_k \partial V_m}. \tag{54}$$

Recall that JF_{ij} is actually the element of the Jacobian matrix of F, which is written as JF. Similarly,

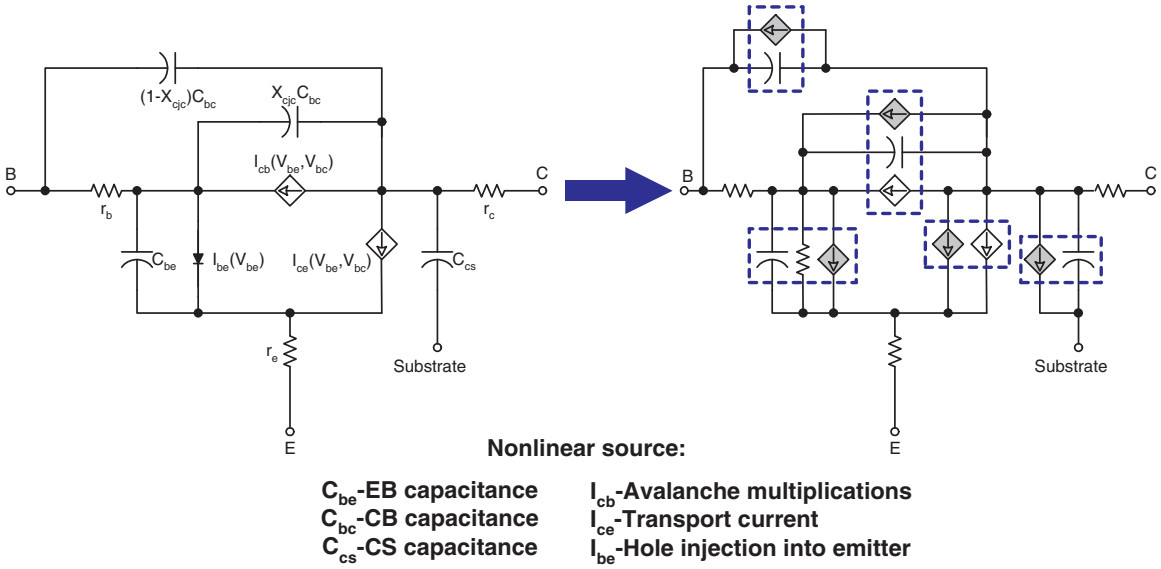

Figure 31: A simplified BJT nonlinear model.

we define *TF1*, *TF2*, and *TF3* as

$$TF1 = \mathrm{JF} = \begin{bmatrix} \mathrm{JF}_{11} & \cdots & \mathrm{JF}_{1n} \\ \vdots & \ddots & \vdots \\ \mathrm{JF}_{n1} & \cdots & \mathrm{JF}_{nn} \end{bmatrix}, \quad TF2 = \left\{ \begin{bmatrix} TF2_{111} & \cdots & TF2_{11n} \\ \vdots & \ddots & \vdots \\ TF2_{1n1} & \cdots & TF2_{1nn} \end{bmatrix} \right. \\ \left. \vdots \right. \\ \left. \begin{bmatrix} TF2_{n11} & \cdots & TF2_{n1n} \\ \vdots & \ddots & \vdots \\ TF2_{nn1} & \cdots & TF2_{nnn} \end{bmatrix} \right\}, \tag{55}$$

and

$$TF3 = \left\{ \begin{bmatrix} TF3_{1111} & \cdots & TF3_{111n} \\ \vdots & \ddots & \vdots \\ TF3_{11n1} & \cdots & TF3_{11nn} \end{bmatrix} \cdots \begin{bmatrix} TF3_{n111} & \cdots & TF3_{n11n} \\ \vdots & \ddots & \vdots \\ TF3_{n1n1} & \cdots & TF3_{n1nn} \end{bmatrix} \right. \\ \left. \vdots \quad \ddots \quad \vdots \right. \\ \left. \begin{bmatrix} TF3_{1n11} & \cdots & TF3_{1n1n} \\ \vdots & \ddots & \vdots \\ TF3_{1nn1} & \cdots & TF3_{1nnn} \end{bmatrix} \cdots \begin{bmatrix} TF3_{nn11} & \cdots & TF3_{nn1n} \\ \vdots & \ddots & \vdots \\ TF3_{nnn1} & \cdots & TF3_{nnnn} \end{bmatrix} \right\} \tag{56}$$

Note that *TF1* is an $n \times n$ matrix, *TF2* is an $n \times n \times n$ tensor product, and *TF3* is an $n \times n \times n \times n$ tensor

product. *TQ1*, *TQ2*, and *TQ3* are defined in a similar manner. These six parameters characterize the nonlinear system.

We abbreviate equations (50) and (51) using the operators $\langle v, v \mid TF2 \rangle$ and $\langle v, v, v \mid TF3 \rangle$ defined in this thesis:

$$\begin{bmatrix} PF2_1 \\ \vdots \\ PF2_n \end{bmatrix} = \langle v, v \mid TF2 \rangle \text{ and } \begin{bmatrix} PF3_1 \\ \vdots \\ PF3_n \end{bmatrix} = \langle v, v, v \mid TF3 \rangle, \tag{57}$$

where $v = [v_1, v_2, \cdots, v_n]^T$. The first operator "multiplies" two n-vectors with one $n \times n \times n$ tensor product and results in a new n-vector, and the second operator "multiplies" three n-vectors with one $n \times n \times n \times n$ tensor product and results in a new n-vector. Equation (52) is then simplified as

$$\begin{aligned} i &= TF1 \times v + \langle v, v \mid TF2 \rangle + \langle v, v, v \mid TF3 \rangle \\ &\quad + \frac{\partial}{\partial t} \left(TQ1 \times v + \langle v, v \mid TQ2 \rangle + \langle v, v, v \mid TQ3 \rangle \right), \end{aligned} \tag{58}$$

where $i = [I_1 - F_{10}, I_2 - F_{20}, \cdots, I_n - F_{n0}]^T$. Similar to the CMNA approach, the i is determined (equal to either 0 or the stimuli) in the linearity calculation. The problem is to solve for v using equation (58). In RF applications, the frequency-domain solution is desired. Therefore, equation (58) is transferred into the frequency-domain using Fourier-series,

$$\begin{aligned} \tilde{i} &= TF1 \times \tilde{v} + \langle \tilde{v}, \tilde{v} \mid TF2 \rangle_\omega + \langle \tilde{v}, \tilde{v}, \tilde{v} \mid TF3 \rangle_\omega \\ &\quad + j\omega \left(TQ1 \times \tilde{v} + \langle \tilde{v}, \tilde{v} \mid TQ2 \rangle_\omega + \langle \tilde{v}, \tilde{v}, \tilde{v} \mid TQ3 \rangle_\omega \right), \end{aligned} \tag{59}$$

where $\tilde{i}$ and $\tilde{v}$ are the Fourier-series of i and v, respectively. $\langle \tilde{v}, \tilde{v} \mid TF2 \rangle_\omega$ is defined as

$$\langle \tilde{v}, \tilde{v} \mid TF2 \rangle_\omega = \begin{bmatrix} \tilde{i}_{CF,21}(\omega) \\ \tilde{i}_{CF,22}(\omega) \\ \vdots \\ \tilde{i}_{CF,2n}(\omega) \end{bmatrix}, \tag{60}$$

where

$$\tilde{i}_{CF,2i}(\omega) = \sum_{(\omega_1,\omega_2)\in A_\omega} \left(\sum_{j=1}^{n} \left(\sum_{k=1}^{n} \left(TF2_{ijk} \times \tilde{v}_j(\omega_1)\tilde{v}_k(\omega_2) \right) \right) \right), i = 1, 2...n, \tag{61}$$

$A_\omega = \{(\omega_1, \omega_2)|\omega_1 + \omega_2 = \omega; \omega_1, \omega_2 \in F_A\}$ is the frequency-pair set. F_A is the frequency set that includes all frequencies in the Fourier-series.

$\langle \tilde{v}, \tilde{v}, \tilde{v} \mid TF3 \rangle_\omega$ is defined as

$$\langle \tilde{v}, \tilde{v}, \tilde{v} \mid TF3 \rangle_\omega = \begin{bmatrix} \tilde{i}_{CF,31}(\omega) \\ \tilde{i}_{CF,32}(\omega) \\ \vdots \\ \tilde{i}_{CF,3n}(\omega) \end{bmatrix}, \tag{62}$$

where

$$\tilde{i}_{CF,3i}(\omega) = \sum_{(\omega_1,\omega_2,\omega_3)\in B_\omega} \left\{ \sum_{j=1}^{n} \left(\sum_{k=1}^{n} \left(\sum_{m=1}^{n} \left(TF3_{ijkm} \times \tilde{v}_j(\omega_1)\tilde{v}_k(\omega_2)\tilde{v}_m(\omega_3) \right) \right) \right) \right\}, \tag{63}$$

$B_\omega = \{(\omega_1, \omega_2, \omega_3)|\omega_1 + \omega_2 + \omega_3 = \omega; \omega_1, \omega_2, \omega_3 \in F_A\}$.

Note that the nonlinear voltage-controlled current sources (shown in Figure 30) are actually the sum-of-products of the voltages at different frequencies. These sources are also referred to as harmonic distortions, where the order of the distortion is equal to the order of the products of the voltages. In small-signal analysis, the higher-order harmonic distortions are negligible. Therefore, only the fundamental and lower-order harmonic distortions need to be considered [58].

In a two-tone analysis, the applied small-signal current sources include two frequency terms ω_1 and ω_2 that satisfy $|\omega_1 - \omega_2| << \omega_1, \omega_2$. At the fundamental frequencies ω_1 and ω_2, the nonlinear-voltage-controlled current sources are negligible, and the equations below are solved to obtain $\tilde{v}(\omega_1)$ and $\tilde{v}(\omega_2)$:

$$\tilde{i}(\omega_1) = (TF1 + j\omega_1 TQ1) \times \tilde{v}(\omega_1), \tag{64}$$

$$\tilde{i}(\omega_2) = (TF1 + j\omega_2 TQ1) \times \tilde{v}(\omega_2). \tag{65}$$

At the frequencies $\omega_1 - \omega_2$ and $\omega_1 + \omega_2$, the second-order harmonic distortion (second-order voltage-controlled current sources) is significant because the applied current source $\tilde{i}$ equals to 0 at these frequencies. However, the third- and higher order harmonic terms are still negligible. The equations below are solved to obtain $\tilde{v}(\omega_1 - \omega_2)$ and $\tilde{v}(\omega_1 + \omega_2)$:

$$0 = \left(TF1 + j(\omega_1 - \omega_2)TQ1 \right) \times \tilde{v}(\omega_1 - \omega_2)$$

$$+ \langle \tilde{v}, \tilde{v} \mid TF2 \rangle_{\omega_1-\omega_2} + j(\omega_1 - \omega_2)\langle \tilde{v}, \tilde{v} \mid TQ2 \rangle_{\omega_1-\omega_2}, \tag{66}$$

$$0 = \Big(TF1 + j(\omega_1 + \omega_2)TQ1 \Big) \times \tilde{v}(\omega_1 + \omega_2)$$
$$+ \langle \tilde{v}, \tilde{v} \mid TF2 \rangle_{\omega_1+\omega_2} + j(\omega_1 + \omega_2)\langle \tilde{v}, \tilde{v} \mid TQ2 \rangle_{\omega_1+\omega_2}. \tag{67}$$

Here, $\tilde{v}(\omega_1)$ and $\tilde{v}(\pm\omega_2)$ that were calculated in equations (64) and (65) are used in the defined operator $\langle \tilde{v}, \tilde{v} \mid TF2 \rangle_{\omega}$.

At the frequencies $2\omega_1 - \omega_2$ and $2\omega_2 - \omega_1$, neither the second- nor the third-order harmonic distortion are negligible. The equations below are solved to obtain $\tilde{v}(2\omega_1 - \omega_2)$ and $\tilde{v}(2\omega_2 - \omega_1)$:

$$0 = \Big(TF1 + j(2\omega_1 - \omega_2)TQ1 \Big) \times \tilde{v}(2\omega_1 - \omega_2)$$
$$+ \langle \tilde{v}, \tilde{v} \mid TF2 \rangle_{2\omega_1-\omega_2} + j(2\omega_1 - \omega_2)\langle \tilde{v}, \tilde{v} \mid TQ2 \rangle_{2\omega_1-\omega_2}$$
$$+ \langle \tilde{v}, \tilde{v}, \tilde{v} \mid TF3 \rangle_{2\omega_1-\omega_2} + j(2\omega_1 - \omega_2)\langle \tilde{v}, \tilde{v}, \tilde{v} \mid TQ3 \rangle_{2\omega_1-\omega_2}, \tag{68}$$

$$0 = \Big(TF1 + j(2\omega_2 - \omega_1)TQ1 \Big) \times \tilde{v}(2\omega_2 - \omega_1)$$
$$+ \langle \tilde{v}, \tilde{v} \mid TF2 \rangle_{2\omega_2-\omega_1} + j(2\omega_2 - \omega_1)\langle \tilde{v}, \tilde{v} \mid TQ2 \rangle_{2\omega_2-\omega_1}$$
$$+ \langle \tilde{v}, \tilde{v}, \tilde{v} \mid TF3 \rangle_{2\omega_2-\omega_1} + j(2\omega_2 - \omega_1)\langle \tilde{v}, \tilde{v}, \tilde{v} \mid TQ3 \rangle_{2\omega_2-\omega_1}. \tag{69}$$

Here, $\tilde{v}(2\omega_1)$, $\tilde{v}(2\omega_2)$, $\tilde{v}(\omega_1 - \omega_2)$, $\tilde{v}(\omega_2 - \omega_1)$, $\tilde{v}(\pm\omega_1)$, and $\tilde{v}(\pm\omega_2)$ are used in the operator $\langle \tilde{v}, \tilde{v} \mid TF2 \rangle_{\omega}$. $\tilde{v}(\pm\omega_1)$ and $\tilde{v}(\pm\omega_2)$ are used in the operator $\langle \tilde{v}, \tilde{v}, \tilde{v} \mid TF3 \rangle_{\omega}$.

Using the approach presented above, the second- and third-order intermodulation-distortion (*IM2* and *IM3*) and input-intercept-power (*IIP2* and *IIP3*) can be analytically calculated. There are no significant numerical errors and convergence problems inherent in this technique because of the small-signal assumption. Moreover, the impact of different device physics on linearity can be easily decoupled: simply set the corresponding $TF2_{jkl}$ and $TF3_{jklm}$ (or $TQ2_{jkl}$ and $TQ3_{jklm}$) to 0. Once the six parameters of the system (*TF1*, *TF2*, *TF3*, *TQ1*, *TQ2*, and *TQ3*) have been determined, the harmonic distortions can be quickly and accurately calculated.

3.4 Small-signal integrated-circuit analyzer

Appling the algorithm described above, we proposed an EDA tool to analyze the broadband noise and linearity in integrated-circuits. The program, which was written in Matlab, adopts the general model and the noise/linearity calculation algorithm discussed in the previous sections. The

principles supporting the device/circuit descriptions in the program are similar to those used in Verilog-A [63]. However, since the simulation algorithm is different, there are some advantages in the software, introduced below:

1. The correlation among different physics noise sources is considered.

2. The linearity calculation is consistent with different input power levels (i.e., no associated numerical errors or convergence problems).

There are also other unique features of the proposed EDA tool, in particular its flexible optimization configuration and device-level analysis capability. Figure 32 shows the framework of this EDA tool. On the device side, the user can turn on or off each individual device physics effect, such as self-heating, impact ionization, etc., or adjust the model parameters of these effects while performing *dc*, *ac*, noise, and linearity simulations. On the circuit side, the user can specify the design target(s), constraint(s) (e.g., gain, *IIP3*, *NF*), and the desired optimization strategy. The tool can be programmed to automatically decrease the design target set and design space to facilitate the optimization process.

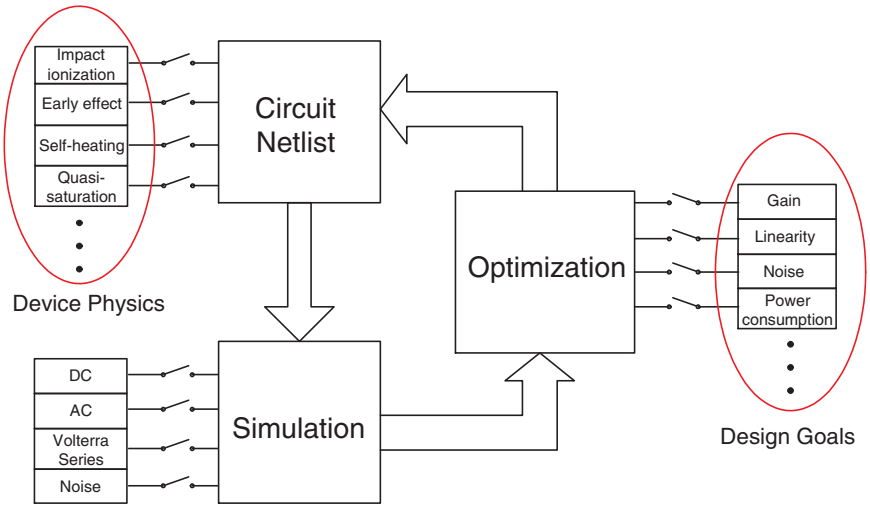

Figure 32: The framework of the proposed EDA tool.

Device-level optimization involves device simulation, device model extraction, circuit simulation, and device engineering. The device structure, the doping level, and the Ge profile are the variables used in the optimization. These variables are the inputs of device simulators such as

MEDICI. The device simulation results (*dc*, *ac*, and noise data) are used to extract the model parameters of the designed device. The extracted model parameters are then applied to a specified circuit design for circuit simulations. The circuit simulation results (gain, *IIP3*, *NF*, etc.) are taken as optimization targets for device design. This technique make it possible to explore and engineer device physics effects to improve circuit performance and provide guidelines for device processing.

In the next chapter, we will demonstrate the application of this software to RF LNA design and optimization, and device-level analysis.

CHAPTER IV

CIRCUIT APPLICATION: RF LNA DESIGN

4.1 Introduction

In an RF receiver, one of the most important features is its noise characteristics, that is, the attribute of collecting the desired input signal while producing less noise. The noise factor F and noise figure NF are the parameters that measure the noise characteristics of RF receivers,

$$F = \frac{SN_{in}}{SN_{out}}, \tag{70}$$

$$NF = 10 \times log_{10}(F), \tag{71}$$

where SN_{in} and SN_{out} are the signal to noise ratios at the input and output of the amplifier, respectively.

Using the definitions above, one can calculate the minimum detectable input power level $P_{in,min}$ of a receiver:

$$P_{in,min} = F \times SN_{req} \times P_{nf}, \tag{72}$$

where SN_{req} is the required signal to noise power ratio at the output of the receiver, and P_{nf} is the input noise floor. $P_{in,min}$ is a fundamental parameter of a receiver: the smaller the $P_{in,min}$, the higher the sensitivity, and the wider area over which the receiver is functional. Note that both SN_{req} and P_{nf} are not determined by the receiver. Therefore, to improve the sensitivity of the receiver (i.e., to reduce $P_{in,min}$), F should be minimized.

In a linear cascade system (Figure 33), the overall noise factor F_{total} can be written as [22]

$$F_{total} = F_1 + \frac{F_2 - 1}{G_1} + \frac{F_3 - 1}{G_1 G_2} \cdots, \tag{73}$$

where F_1, F_2, and F_3 are the noise factors of stages 1, 2, and 3, G_1, G_2, and G_3 are the gains of stages 1, 2, and 3, respectively. Observe that if the gain of the first stage (G_1) is high enough, the

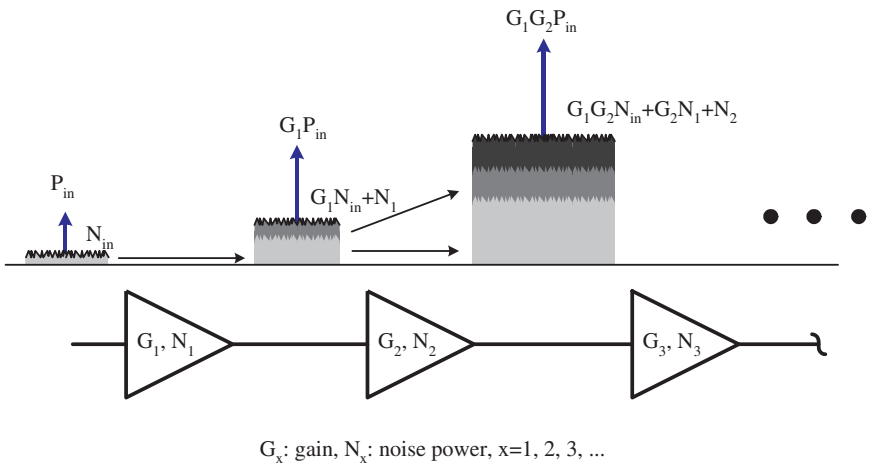

Figure 33: Signal and noise gain in cascade amplifiers.

overall noise factor is dominated by the noise factor of the first stage (F_1). Therefore, a low-noise-amplifier (LNA), which provides a decent gain with a smaller F, is required and must be attached in front of the receiver to obtain a low F_{total}.

Compared with other RF building blocks such as mixers and oscillators, the circuit structure of an LNA is simple since the addition of devices produces more noise. There are three main circuit topologies used for LNAs: single-transistor, cascade, and cascode [64]-[66]. Figure 34 shows the three different LNA equivalent circuit structures. Compared with single-transistor LNAs, the cascade and cascode LNAs provide higher gain and isolation. However, the single-transistor LNAs require less *dc* power consumption and can achieve a lower noise factor. The most appropriate topology is selected based on the application requirement.

From a device perspective, the SiGe HBT is a better contender for Si-based RF LNAs because of its high speed, high gain, low noise, and good linearity [4]. Improvements in the device technology have broadened the design space for LNAs. On the other hand, it also demands higher levels of optimization for the LNAs to achieve better performance, indicating that the design and optimization is not only at the circuit level (e.g., with different circuit structures and biasing currents), but also at the device level (e.g., with different device geometries and device profiles). In this chapter, we will focus on general design concerns and optimization strategies for LNAs. An example of an inductively-degenerated cascode SiGe LNA optimization will then be demonstrated.

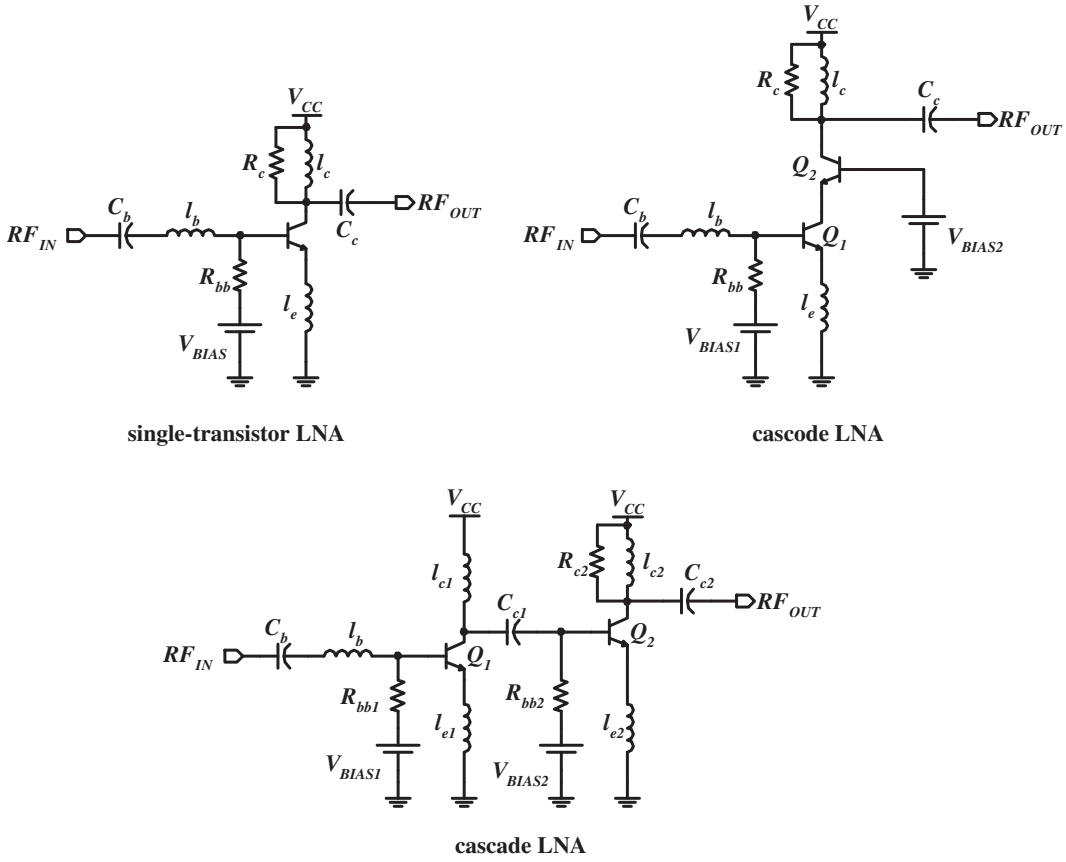

Figure 34: The three different LNA equivalent circuit structures.

4.2 RF LNA design concerns

4.2.1 Noise factor/figure

According to linear two-port noise theory [67], the noise factor (F) or noise figure (NF) is a function of the source impedance (Y_S) and the noise parameters of the two-port system (F_{min}, R_n, and Y_{opt}):

$$F = F_{min} + \frac{R_n}{G_S} \left| Y_S - Y_{opt} \right|^2, \tag{74}$$

where G_S is the real part of Y_S, F_{min} is the minimum noise factor, R_n is the noise impedance, and Y_{opt} is the optimum admittance of the system. The noise parameters are functions of the biasing current and device properties (both active and passive). Figure 35 shows the noise factor and minimum noise factor as a function of the collector current (I_C) of a single-transistor SiGe HBT LNA.

In RF LNAs, the source impedance is fixed, usually at 50 Ω. Then, the optimization goal is to

engineer the noise parameters of the system to obtain a lower F. One method is to determine the minimum F_{min} in design space and then tune the Y_{opt} (without changing F_{min}) to be close to Y_S. This approach produces the lowest F for the amplifier. As shown in Figure 35, for example, the minimum F_{min} is at $I_C = 3.0$ mA. Therefore, to obtain the lowest NF (i.e., 0.93 dB), I_C should equal 3 mA and $Y_{opt} = Y_S$.

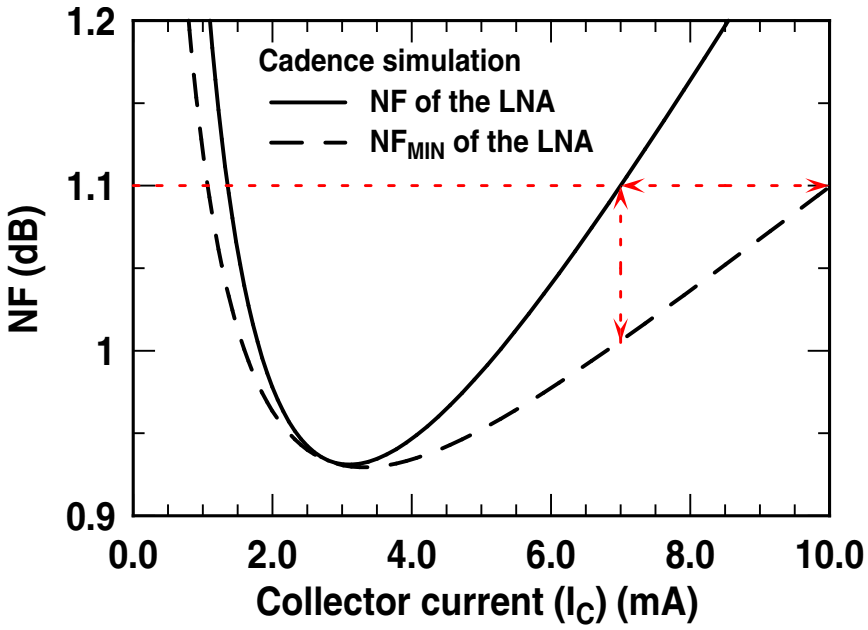

Figure 35: The actual noise factor and minimum noise factor as a function of the collector current (I_C) of a single-transistor SiGe HBT LNA.

In reality, however, other specifications such as gain and linearity of the LNA also need to be considered, and trade-offs between the noise factor and these specifications are unavoidable. In this case, the noise figure F does not need to be tuned to F_{min} because F_{min} also depends on design variables, particularly the device geometry and biasing current [68]. As shown in Figure 35, the actual NF at $I_C = 6$ mA is lower than NF_{min} at $I_C = 10$ mA. This implies that relaxing the noise factor requirement can either be traded for a larger biasing current range while $Y_{opt} = Y_S$ holds, or a smaller I_C range but more flexible noise matching requirement ($Y_{opt} \neq Y_S$), hence permitting a larger device geometry range. Due to these concerns, general optimization strategies will be developed in the next section.

4.2.2 Gain and matching

The gain of LNA also affect the overall noise factor of the system. Among various gain definitions, the transducer gain (G_T) and available gain (G_A) are used as the figures-of-merit since the source impedance of LNA is fixed.

The transducer gain is the ratio of the actual power delivered by the amplifier to an arbitrary load to the power available from the source. It is a function of the s-parameters of the amplifier (s_{11}, s_{12}, s_{21}, and s_{22}) and the source and load reflection coefficients (Γ_S and Γ_L)

$$G_T = |s_{21}|^2 \frac{\left(1 - |\Gamma_S|^2\right)\left(1 - |\Gamma_L|^2\right)}{|1 - \Gamma_S s_{11}|^2 \, |1 - \Gamma_L \Gamma_2|^2}, \tag{75}$$

where

$$\Gamma_2 = s_{22} + \frac{s_{12} s_{21} \Gamma_S}{1 - s_{11} \Gamma_S}. \tag{76}$$

The available gain is the ratio of the power available at the output of the amplifier to the power available from the source. It is actually the maximum transducer gain (obtained by sweeping the load impedance) and is a function of the s-parameters and the source reflection coefficients

$$G_A = |s_{21}|^2 \frac{1 - |\Gamma_S|^2}{|1 - \Gamma_S s_{11}|^2 \left(1 - |\Gamma_2|^2\right)}. \tag{77}$$

The source impedance is usually 50 Ω. Therefore, $G_A = |s_{21}|^2/(1 - |s_{22}|^2)$.

When an LNA is designed exclusively as an RF building block, low voltage-standing-wave ratios (VSWRs) at the input and output of the LNA are also required. This requirement offers good matching between the LNA and other blocks (e.g., antenna, mixer, filter). Therefore, the $|s_{11}|$ and $|s_{22}|$ of the LNA should be minimized. In ideal conditions, $s_{11} = s_{22} = 0$ and $G_A = |s_{21}|^2$. Thus, the optimization goals concerning the gain and matching are to maximize $|s_{21}|$ and minimize $|s_{11}|$ and $|s_{22}|$.

4.2.3 Linearity

Another concern in LNAs is linearity, which confines their maximum input power. For circuits such as LNAs operating in the small-signal range, the figure-of-merit of linearity is the input third-order intercept point (*IIP3*).

Figure 36 shows the first- and third-order output power as a function of input power of the LNAs with different *IIP3*s. In an RF receiver, the third-order distortion of the output power is an unexpected signal, and should be lower than a required level (e.g., the noise floor) to maintain the output signal quality. Correspondingly, the input power should be lower than $P_{in,max}$,

$$P_{in,max} = IIP3 - (IIP3 + gain - P_{nf})/3, \tag{78}$$

where P_{nf} is power level of the noise floor.

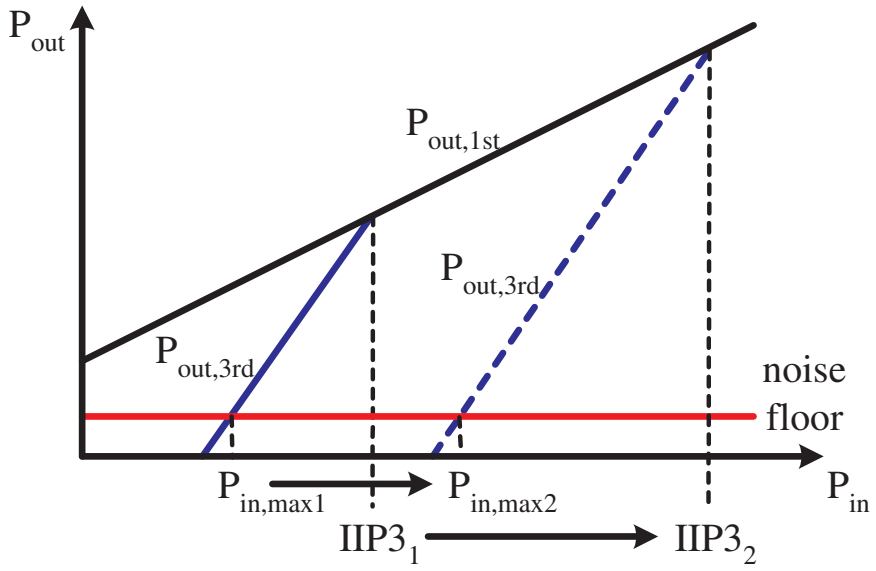

Figure 36: The first- and third-order output power as a function of input power of the LNAs with different *IIP3*s.

Observe that the LNA with higher *IIP3* exhibits a higher $P_{in,max}$, indicating a wider range for acceptable input power. Therefore, maximizing *IIP3* is another goal of LNA optimization.

4.2.4 Power consumption and stability

The total *dc* power is approximately equal to the product of the biasing current and supply voltage:

$$P_{DC} = \sum_i I_{bias,i} V_{bias,i}, \tag{79}$$

where $I_{bias,i}$ and $V_{bias,i}$ are the *dc* injected current and voltage supply at pad i, respectively. The supply voltage is typically fixed, therefore a smaller biasing current is desirable in low power LNA

designs. Additionally, a robust LNA should be unconditionally stable over any frequency range. These requirements are generally considered as design constraints rather than optimization goals.

4.3 SiGe cascode LNA design

4.3.1 Optimization strategies

Without loss of generality, the optimization goals for RF LNAs are low noise (*NF*), high gain ($|s_{21}|^2$), low VSWR ($|s_{11}|$ and $|s_{22}|$), and high linearity (*IIP3*) with the low power consumption and stability constraints, and the design variables are the values of inductors (L_x), capacitors (C_y), and resistors (R_z), the device geometry (emitter length L_E),[1] and biasing current (I_C).[2] The optimization goals can be written as functions of design variables (e.g., $NF(L_x, C_y, R_z, L_E, I_C)$, $IIP3(L_x, C_y, R_z, L_E, I_C)$). Therefore, the problem can be generalized as optimizing a group of goal-functions in an N-dimensional space, where N is the number of the design variables.

For a better evaluation of the overall performance, we first choose the goal-functions that will not be considered in the trade-off (e.g., $|s_{11}|$ and $|s_{22}|$). The dimension of the design space decreases as these goal-functions are optimized. In inductively-degenerated cascode LNAs (Figure 34), for example, the values of inductors and capacitors are determined when $|s_{11}|$ and $|s_{22}|$ are minimized, and consequently the design space is reduced to 2-D (L_E and I_C).

Second, the values of those goal-functions that are of interest to us (e.g., *NF*, gain, and *IIP3*), are calculated in the design sub-space, and the contours of these functions are drawn. In this example, the contour-lines of *NF*, gain, and *IIP3* are plotted in the 2-D ($L_E - I_C$) space.

Moreover, the design sub-space shrinks when the additional design constraints are considered. For instance, the power consumption constraint requires that I_C should be smaller than a required level ($I_{C,max}$). Therefore, the design $L_E - I_C$ sub-space diminishes from $(0, \infty)\times(0, \infty)$ to $(0, \infty)\times(0, I_{C,max})$. The optimum design point, as shown later, can be easily found in this contour plot on the design sub-space.

[1] For brevity, the effect of device geometry is viewed as the effect of laying out multiple unit devices in parallel. Here, the emitter length L_E is used and is equivalent to a device scaling factor.

[2] I_C denotes the collector biasing current. For MOSFETs, I_{DS} is used to denote the drain-source biasing current

4.3.2 Device models

In this chapter, a simplified device model is used in analysis to derive analytical expressions and design rules of thumb. In actual design experience, delicate compact models such as VBIC, HICUM and MEXTRAM are used so that the parasitics and second-order effects are also considered.

IBM 5HP SiGe HBTs are used in this design. Table 2 summarizes the measured transistor parameters of a typical SiGe HBT unit. Figure 31 shows the large-signal nonlinear model of this SiGe HBT. Here, I_{CE} represents the collector current transported from the emitter, I_{BE} represents the hole injection into the emitter, and I_{CB} represents the avalanche multiplication current

$$I_{CB} = I_{CE}\cdot(M-1) = I_{C0}(V_{BE})\cdot F_{Early}\cdot(M-1), \tag{80}$$

where $I_{C0}(V_{BE})$ is I_C measured at zero V_{CB}, M is the avalanche multiplication factor, and F_{Early} is the Early effect factor. M-1 and F_{Early} were experimentally extracted using a technique proposed in [69].

Table 2: Measured parameters of a SiGe HBT. The device size is $0.5 \times 20 \times 2\ \mu m^2$.

Peak β	110
Peak f_T	51 GHz
τ_{ec}	2.7 ps
$r_{bb}@I_C = 10$ mA	8.9 Ω
BV_{ceo}	3.3 V

M-1 is often modeled only as a function of V_{CB}. In the SiGe HBTs used in this work, however, M-1 is also a strong function of J_C [36]. The following equation was used to describe the V_{CB} and J_C dependence of M-1

$$M-1 = \frac{V_{CB}}{aV_{CB0}},\cdot \exp\left(-\frac{V_m^b}{a^{1-b}\cdot V_{CB}^b}\right), \tag{81}$$

where a is an empirical coefficient which takes into account the Kirk effect, and

$$a = 1 - \tanh\left[\frac{I_C}{I_{C0}}\exp\left(\frac{V_{CB}}{V_R}\right)\right], \tag{82}$$

where V_{CB0}, V_m, V_R, I_{C0} and b are model parameters, and b is typically $0.33-0.5$. These parameters must be determined by fitting the measured M-1 data. Figure 37 shows the measured and model-fitted M-1 as a function of the collector current density J_C at different V_{CB}. The model agrees well

with the measured data.

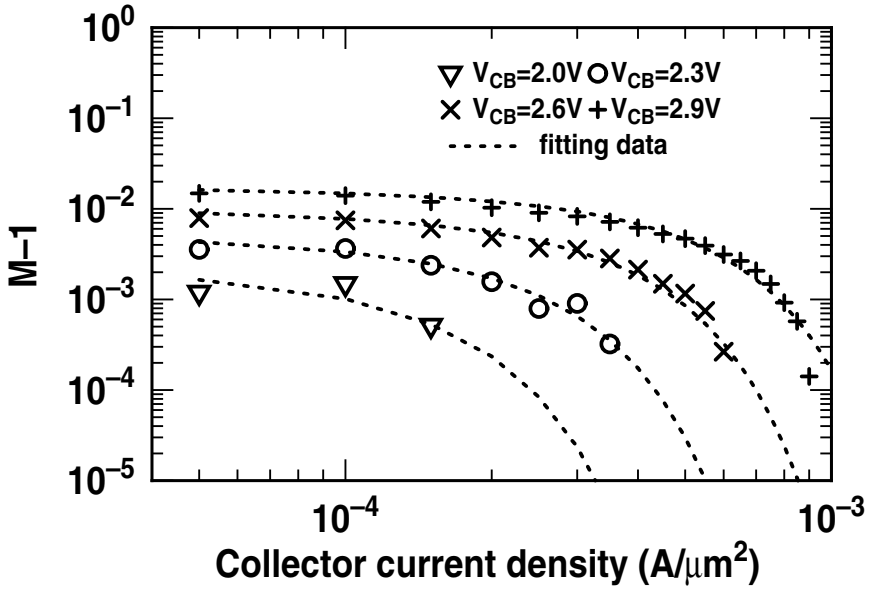

Figure 37: The measured and model-fitted M-1 as a function of the collector current J_C at different V_{CB}.

In addition, C_{cb} is the collector base junction depletion capacitance, C_{cs} is the collector substrate depletion capacitance, and C_{be} consists of diffusion capacitance C_{de} and depletion capacitance C_{te} according to

$$C_{be} = C_{de} + C_{te}, \tag{83}$$

$$C_{de} = g_m \cdot \tau_f. \tag{84}$$

In this model, six nonlinear sources are considered: nonlinear transconductance (g_m), nonlinear EB conductance (g_{be}), three nonlinear capacitances (C_{be}, C_{bc}, and C_{cs}), and avalanche multiplication (I_{cb}).

In the RF LNA noise simulation, only broadband noise is considered.[1] Figure 29 shows the noise model of the SiGe HBT. The noise sources are expressed in equations (38)-(41).

Before the analysis, the presented linearity calculation algorithm in the EDA tool is further verified by simulation results obtained using HP-ADS. Figure 38 shows the *IIP3* versus biasing I_C for a single transistor amplifier obtained using our Volterra-series approach and HP-ADS. At

[1] Since the devices in LNA are operating in the small-signal range, the device-induced low frequency noise is negligible at radio frequencies

input powers below −20 dBm and higher than −50 dBm, the deviation between the *IIP3*s is less than 0.05 dB, demonstrating the good precision offered by this approach. Using this EDA tool, we

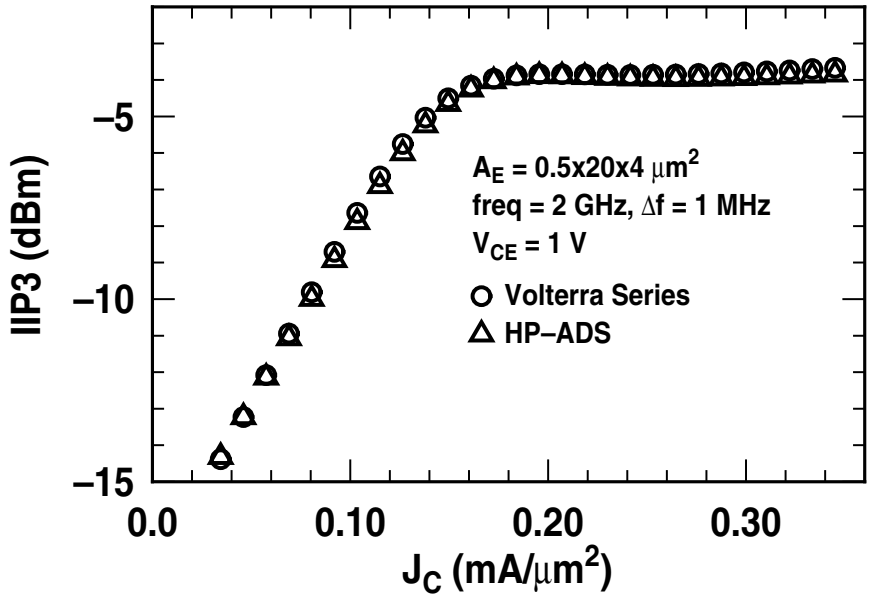

Figure 38: *IIP3* as a function of bias current in a single transistor. $A_E = 0.5 \times 20 \times 2\ \mu m^2$, $V_{ce} = 1$ V. The tone spacing is 1 MHz.

can now perform a systematic analysis and optimization in a SiGe inductively-degenerated cascode LNA.

4.3.3 I_C dependence analysis

The simulation results show that *IIP3* does not increase monotonically as collector current increases when inductors l_e and l_b are adjusted for input-impedance matching. Figure 39 shows *IIP3* as a function of collector current for four different emitter lengths. Peaks of *IIP3* can be observed for each emitter length. For instance, an *IIP3* peak occurs at $I_{ce} \simeq 6$ mA when the emitter length $= 40\ \mu m$. Past this point, a further increase in collector current degrades *IIP3*.

To understand the physical origin of the peak, *IIP3* was simulated by turning on the individual physical nonlinearities as well as turning on all the nonlinearities. Figure 40 shows the resultant individual and overall *IIP3* as a function of collector current for an emitter length of 40 μm. Observe that when $I_{ce} = 6-8$ mA, the overall *IIP3* (∇) is higher than *IIP3* caused by either g_m ($\times$), g_{be} (o), or C_{be} (+), indicating that there exists finite nonlinearity cancellation in these devices. Note that there

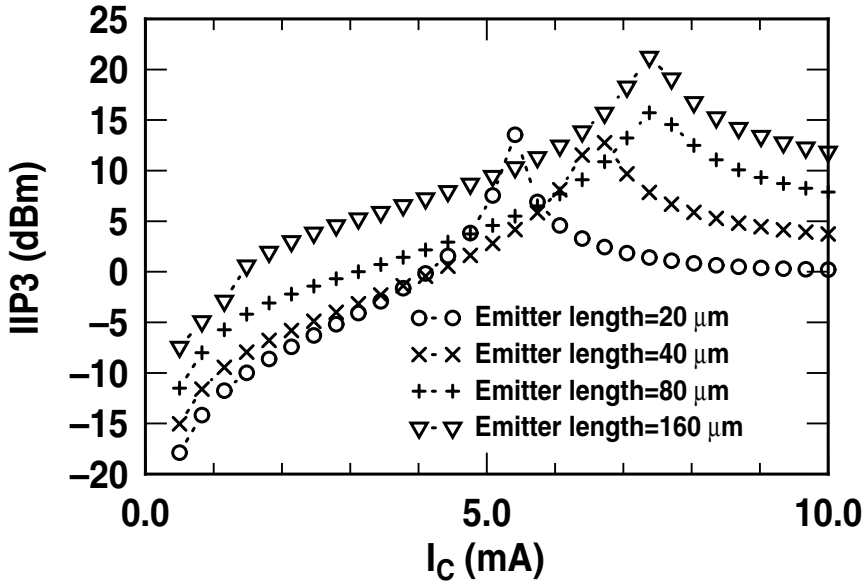

Figure 39: *IIP3* as a function of I_C in 0.5 × 20, 0.5 × 40, 0.5 × 80, and 0.5 × 160 μm^2 transistors at 2 GHz.

is a peak for the total *IIP3*, but no peak for *IIP3* caused by any individual nonlinearity, implying that the cancellation is maximized at the *IIP3* peak. Increasing collector current has two effects on *IIP3*: 1) It monotonically decreases the impedance of the EB junction, thereby reducing the first-order (linear) *ac* voltage on the EB junction, which controls the g_m, g_{be}, and C_{be} nonlinearities. As a result, the *IIP3* calculated by turning on these individual nonlinearities monotonically increases with increasing I_C; 2) It changes the cancellation among individual nonlinearities, which is maximized at a certain collector current. The resulting increase of *IIP3* leads to a peak in the overall *IIP3* obtained by turning on all of these nonlinearities. The overall *IIP3* can be maximized at each collector current. The distortions generated by each individual nonlinear source cancel out in part, which makes the overall distortion lower than the individual distortion.

Note that the observed non-monotonic behaviour of *IIP3* is in direct contrast to conventional wisdom that *IIP3* improves monotonically with increasing collector current [22]. The underlying reason is that l_e and l_b are adjusted for input impedance matching as the collector current increases. The cancellation of *IIP3* dominates, producing an *IIP3* peak as shown in Figure 40.

Figure 41 shows gain as a function of collector current for four different emitter lengths. Observe that the gain increases monotonically as collector current increases. Physically, it increases

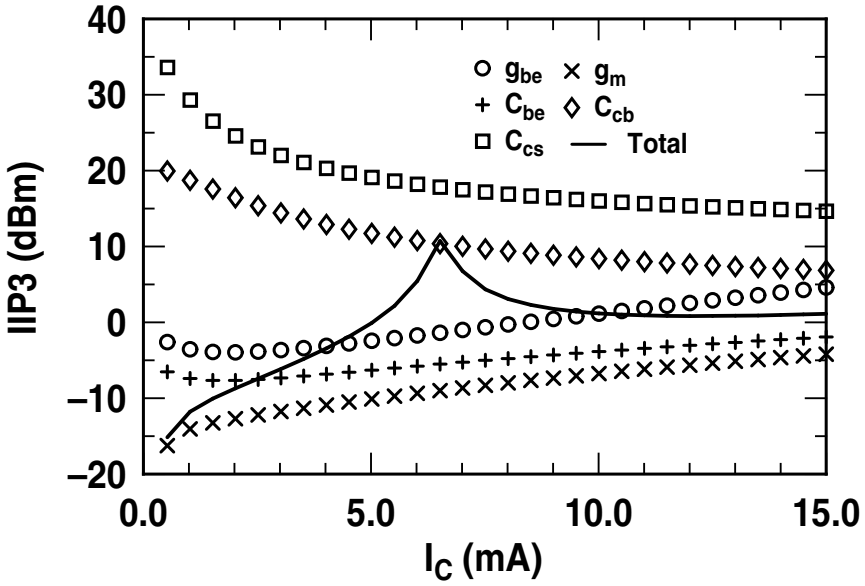

Figure 40: Comparison of collector current dependence of *IIP3* caused by different nonlinear sources at 2 GHz, for an emitter length = 40 μm.

gradually at high collector current because the diffusion capacitance (C_{de}) across the EB junction increases as collector current increases and thus dominates the overall conductance of the EB junction.

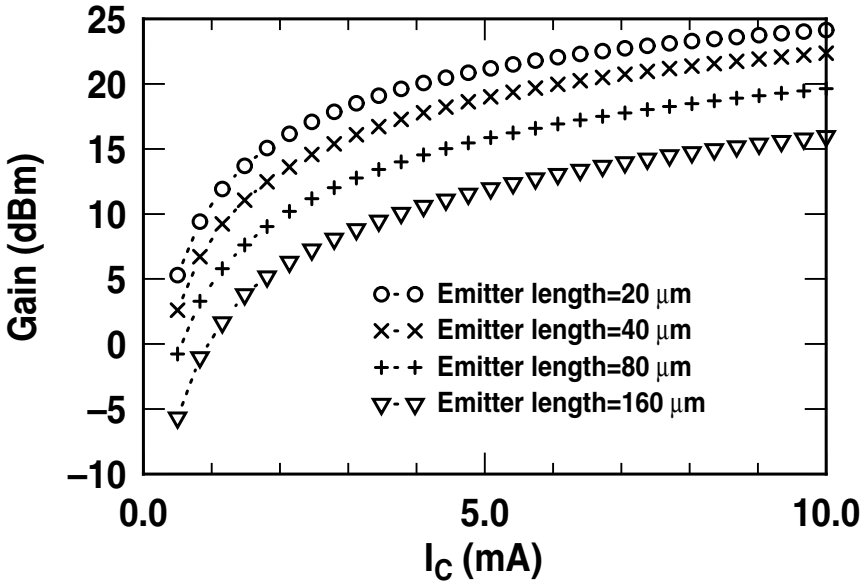

Figure 41: Gain as a function of I_C in 0.5 × 20, 0.5 × 40, 0.5 × 80, and 0.5 × 160 μm^2 transistors at 2 GHz.

Figure 42 shows *NF* as a function of collector current for four different emitter lengths. As expected, the collector current shot noise is dominant at lower I_C, *NF* increases as I_C decreases because of the gain decreases. At higher I_C, the base current shot noise, which is proportional to I_C, dominates *NF*. Thus, there is an optimum value of I_C that minimizes *NF*, as expected.

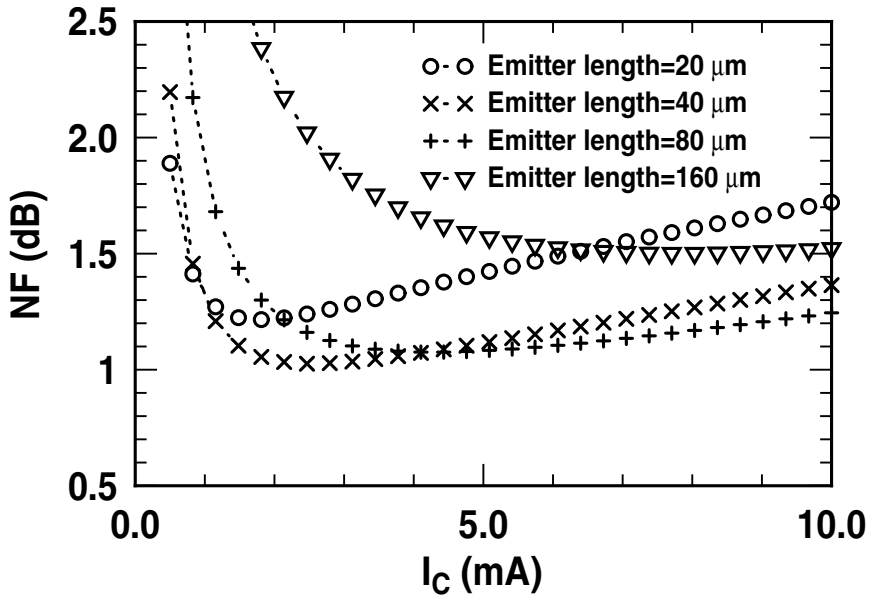

Figure 42: *NF* as a function of I_C in 0.5 × 20, 0.5 × 40, 0.5 × 80, and 0.5 × 160 μm^2 transistors at 2 GHz.

4.3.4 Geometry dependence analysis

Figure 43 shows *IIP3* as a function of emitter length for different values of I_C at 2 GHz. At $I_C = 6\ mA$, a peak in *IIP3* can be observed at an emitter length of 30 μm. At $I_C = 4\ mA$, the *IIP3* peak occurs at emitter lengths lower than 10 μm. The peak is due to nonlinearity cancellation. However, at $I_C = 7.5\ mA$, *IIP3* increases monotonically as the emitter length increases up to 100 μm. This is because the cancellation is a strong function of collector current and is not dominant at $I_C = 7.5\ mA$, as shown below.

Figure 44 shows the resultant individual and overall *IIP3* as a function of emitter length with $I_C = 6\ mA$ at 2 GHz. The overall *IIP3* increases up to its peak value at an emitter length of about 33 μm, then rolls off to a nearly constant value as the length increases beyond 33 μm. The emitter length has two effects on *IIP3*: 1) Increasing the emitter length with a fixed bias current

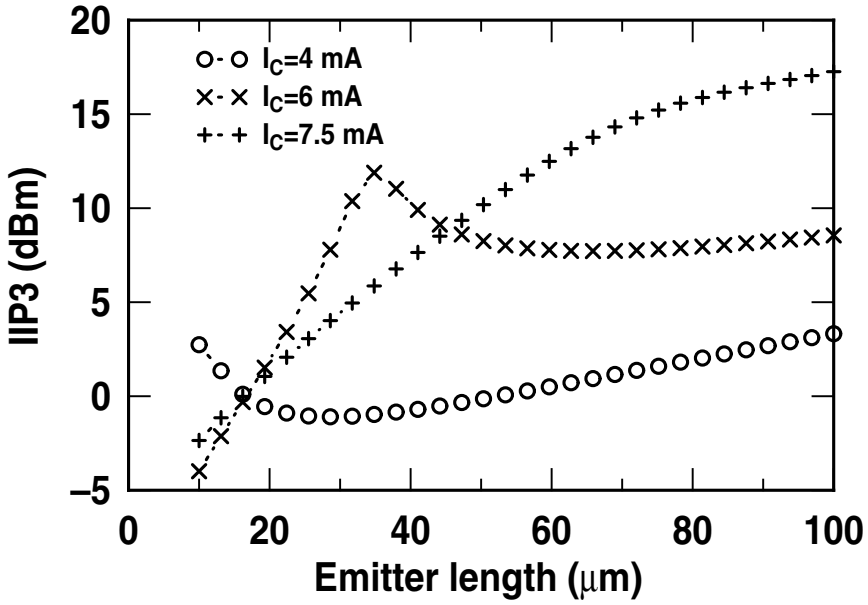

Figure 43: *IIP3* as a function of emitter length, with $I_C = 4, 6, 7.5$ mA at 2 GHz.

increases the EB depletion capacitance, thus decreasing the magnitude of the impedance across the EB junction. Because the input impedance is fixed at 50 Ω, the *ac* current from the source injected into the EB junction does not change as emitter length increases. Thus the first-order *ac* voltage on the EB junction decreases, consequently, the *IIP3* calculated by turning on these individual nonlinearities increases monotonically; 2) It changes the cancellation between individual nonlinearities. For $I_C = 6\ mA$, the cancellation dominates the distortion and is maximized at an emitter length of 30 μm, and thus a peak in *IIP3* occurs. For $I_C = 7.5\ mA$, other effects dominate the distortion and no *IIP3* peak occurs.

Figure 45 shows the gain as a function of emitter length with different values of I_C. The depletion capacitance increases as the emitter length increases, decreasing the gain.

Figure 46 shows noise figure as a function of emitter length with different values of I_C at 2 GHz. For short emitter lengths, the thermal noise caused by the base resistance is dominant. For long emitter lengths, however, the collector current shot noise contribution dominates *NF*. Therefore, there is an optimum emitter length that balances the noise caused by r_b and I_C.

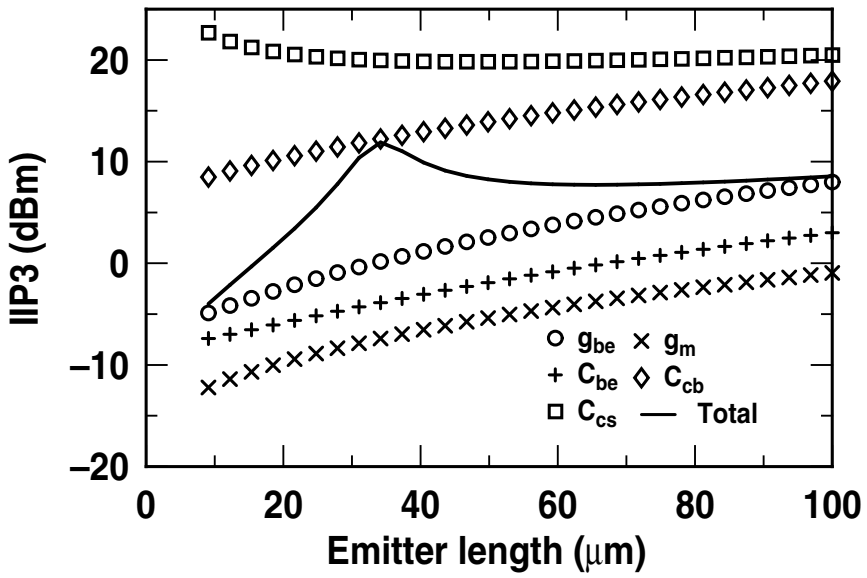

Figure 44: Comparison of emitter length dependence of *IIP3* caused by different nonlinear sources at 2 GHz, for $I_C = 6$ mA.

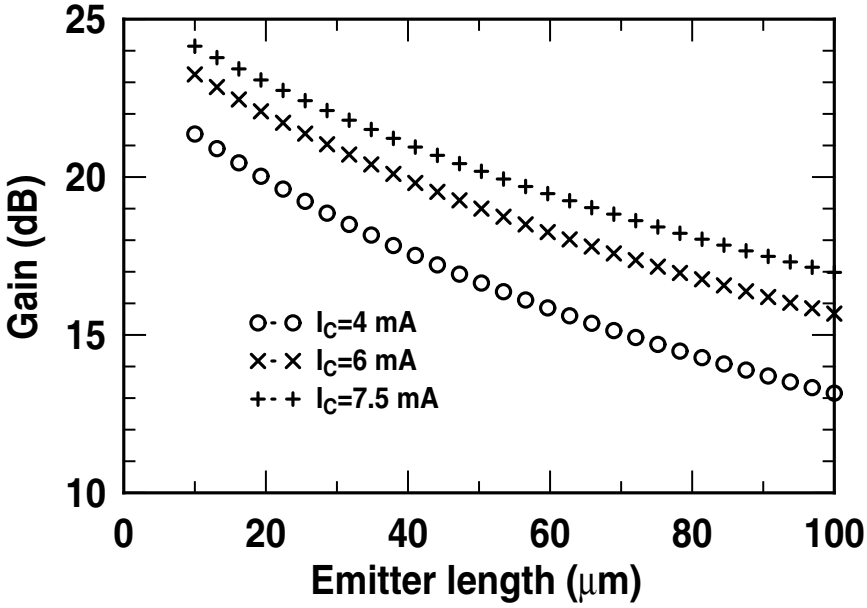

Figure 45: Gain as a function of emitter length, with $I_C = 4, 6, 7.5$ mA at 2 GHz.

4.3.5 Optimum LNA design

A determination of the optimum emitter length and I_C which balances *IIP3*, gain, and noise figure can be obtained by an overall consideration of these simulation results. Figure 47 shows *IIP3* contours as a function of emitter length and collector current. The design space for $NF \leq 1.2$ *dB*

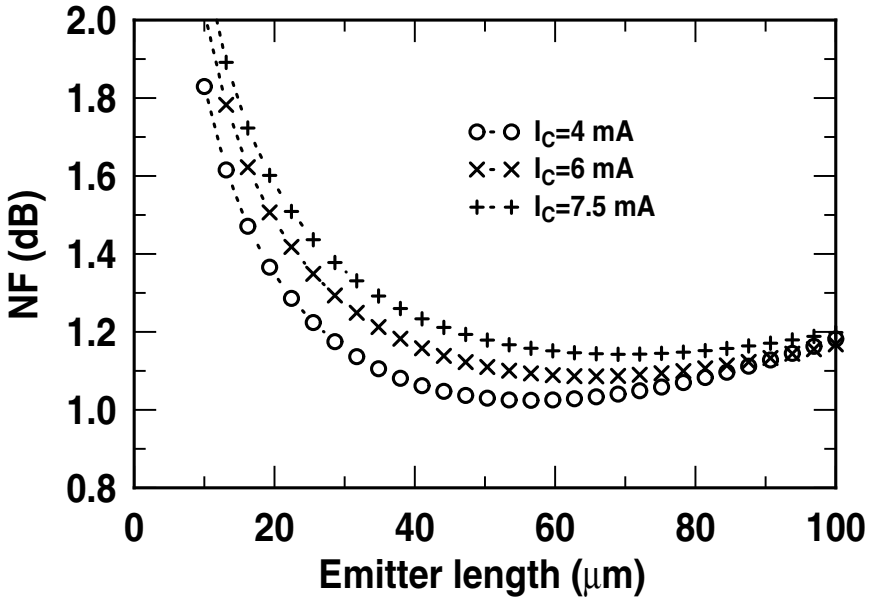

Figure 46: *NF* as a function of emitter length, with $I_C = 4, 6, 7.5$ mA at 2 GHz.

is within the dashed line, and the design space for gain $\geq$ 15 *dB* is above the dash-dotted line. Within the design space that meets these criteria ($NF \leq 1.2$ *dB* and gain $\geq$ 15 *dB*), observe that *IIP3* changes dramatically, from approximately -5 dBm to 15 dBm. The optimum design point for a maximum *IIP3* is thus an emitter length = 80 μm and $I_C = 7.5$ mA. The maximum *IIP3* is then above 15 dBm, with a resultant noise figure of 1.15 dB.

In contrast, *IIP3* at the design point optimum for noise figure ($I_C = 4$ mA, emitter length = 60 μm) is only 0 dBm. The LNA design point optimum for *IIP3* is a better overall choice, because the noise figure is still near its minimum, while *IIP3* is significantly higher (by 15 dB). The disadvantage of this approach, however, is that the required bias current is 3.5 mA higher.

Therefore, for an optimized design (80 μm and 7.5 mA), *IIP3* = 15.8 dBm, gain is 18 dB, *NF* = 1.15 dB, and $|s_{11}|, |s_{22}| < -30$ dB. Furthermore, if the power consumption constraint shrinks to $I_C \leq 5.5$ mA, an *IIP3* of 5 dBm can be obtained at $L_E = 50$ μm and $I_C = 5.5$ mA, with a near-minimum noise figure of 1.08 dB.

4.4 Analytical expressions and design rules-of-thumb

The procedure introduced above is a numerical methodology that produces accurate simulation results. Furthermore, analytical expressions of the values of required inductors, gain, *NF*, and *IIP3*

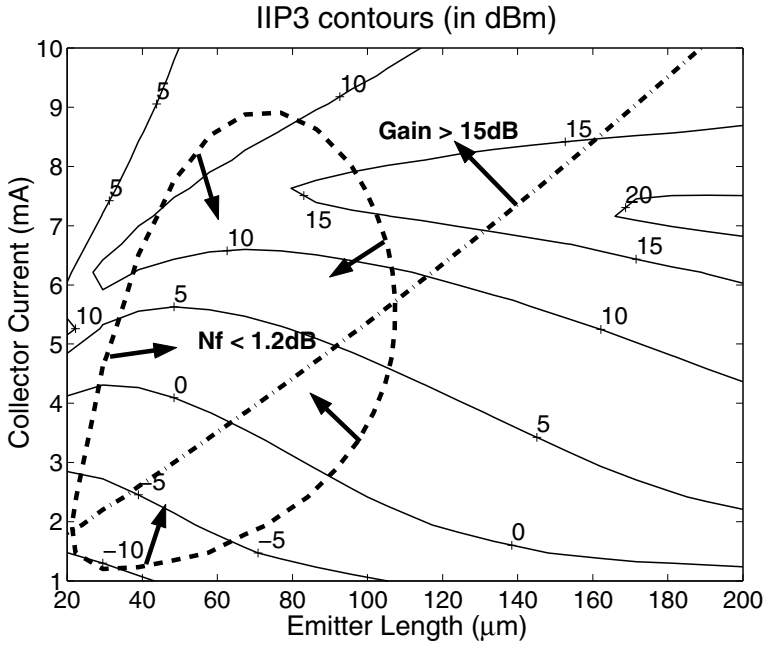

Figure 47: *IIP3* contours as a function of emitter length and collector current for an input impedance matched LNA at 2 GHz. The dashed line is the noise figure contour at 1.2 dB, and the dash-dot line is the gain contour at 15 dB.

can also be derived by simplifying the device model [8]. These analytical expressions provide the design rules of thumb.

To derive the analytical expressions, we neglect the emitter and collector resistance, base-collector and collector-substrate capacitance, and avalanche multiplication current. The analytical expressions of l_e and l_b can be written as

$$l_e \simeq \frac{R_s}{2\pi f_T} - \frac{2\pi f_T(1/g_{be} - R_s)}{\omega^2\beta^2}, \tag{85}$$

$$\begin{aligned} l_b &\simeq \frac{2\pi f_T(1/g_{be} - R_s)}{\omega^2\beta} - l_e, \\ &\simeq \frac{2\pi f_T(1/g_{be} - R_s)}{\omega^2\beta} - \frac{R_s}{2\pi f_T}, \end{aligned} \tag{86}$$

where $\omega = 2\pi f$ is the circuit operating frequency, $R_s = 50\ \Omega$ is the source impedance, and f_T is the cut-off frequency at the chosen $L_E - I_C$ (emitter length−bias current) point. Note that g_{be} was neglected in the traditional input impedance matching equations, assuming $1 << \omega\beta/2\pi f_T$ and $R_s << 1/g_{be}$ [70]. However, the assumption of $2\pi f_T << \omega\beta$ is no longer valid when f_T is much

higher than the operational frequencies for SiGe HBT LNAs. In RFIC fabrication, smaller l_b[1] is preferred because the Q-factor of typical on-wafer inductors is small, and hence will degrade noise performance. From the equations above, when I_C is fixed, the required l_b becomes smaller as f_T decreases. At low injection, f_T decreases as the current density ($J_C = I_C/W_E/L_E$, where W_E is the emitter width) decreases, so a larger L_E requires a smaller l_b.

Assuming that $\omega << 2\pi f_T$, the current gain of the common-base stage is close to unity. Moreover, since the input impedance is matched to the source resistance R_s, the current injected into the EB junction is constant. Thus, the current gain equals to

$$\frac{\beta g_{be}}{g_{be} + j\omega C_{be}} = \frac{\beta}{1 + j\omega\beta/2\pi f_T}, \tag{87}$$

and the gain of power can be written as

$$gain = \frac{\beta^2 R_{load}}{[1 + (\omega\beta/2\pi f_T)]^2 R_s}. \tag{88}$$

At low injection, as J_C increases, f_T increases, and thus the gain increases. Hence gain increases with increasing I_C and fixed L_E, and decreases with increasing L_E and fixed I_C, as expected.

The noise figure can be written as

$$\begin{aligned}
NF &= 10\log_{10}(1 + n_{ib} + n_{ic} + n_{vb}), && (89)\\
n_{ib} &= \frac{(g_{be}R_s)^2 + [B(1 - g_{be}R_s)]^2}{2g_{be}R_s}, && (90)\\
n_{ic} &= \frac{4(g_{be}R_s)^2 + [g_{be}R_s/B + B(1 - g_{be}R_s)]^2}{2g_m R_s}, && (91)\\
n_{vb} &= \frac{r_b}{R_s}, && (92)
\end{aligned}$$

where $B = 2\pi f_T/\omega\beta$. Under an overall power consumption constraint, I_C normally is less than 10 mA, meaning $g_{be}R_s << 1$. Rewriting the equation above, one obtains

$$\begin{aligned}
n_{ib} &= \frac{(g_{be}R_s)^2 + B^2}{2g_{be}R_s}, && (93)\\
n_{ic} &= \frac{4(g_{be}R_s)^2 + [g_{be}R_s/B + B]^2}{2g_m R_s}, && (94)\\
n_{vb} &= \frac{r_b}{R_s}. && (95)
\end{aligned}$$

[1]The value l_e is usually much smaller than l_b, thus the effect of the parasitics of l_e on noise is negligible.

According to these equations, when L_E is fixed, at lower I_C, the collector current shot noise (n_{ic}) is dominant. *NF* increases as I_C decreases. At higher I_C, the base current shot noise (n_{ib}), which is proportional to I_C, dominates *NF*. Thus, there is an optimum value of I_C that minimizes *NF*.

When I_C is fixed, at a small L_E, the thermal noise caused by the base resistance (n_{rb}) is dominant. At a large L_E, however, the collector current shot noise (n_{ic}) contribution dominates *NF*. Therefore, there is an optimum L_E that balances the noise caused by r_b and I_C.

Applying the Volterra series, one can derive (see Appendix A for details)

$$IM3 \simeq \left| \frac{3}{4} \cdot C(\omega_1, \omega_2) \cdot L(\omega) \cdot (1 - G(2\omega_1) - 2G(\omega_1 - \omega_2)) \right|, \tag{96}$$

where

$$\begin{aligned} C(\omega_1, \omega_2) &= \frac{1}{6V_t^2} \cdot v_{in,1}(\omega_1) \cdot v_{in,1}(-\omega_2) \\ &\simeq \frac{1}{6V_t^2} \cdot |v_{in,1}^2(\omega_1)|, \end{aligned} \tag{97}$$

and $v_{in,1}$ is the 1st order *ac* voltage on the EB junction. In addition,

$$L(\omega) = \frac{V_t \cdot K(\omega) \cdot G(\omega)}{I_C}, \tag{98}$$

where

$$G(\omega) = \frac{A(\omega) \cdot I_C}{B(\omega) + A(\omega) \cdot I_C}, \tag{99}$$

$$K(\omega) = \frac{B(\omega)}{A(\omega) \cdot V_t}, \tag{100}$$

$$A(\omega) = \left(j\omega \cdot \tau_f + \frac{1}{\beta} \right) \cdot (Z_b(\omega) + Z_e(\omega)) + Z_e(\omega), \tag{101}$$

$$B(\omega) = V_t \cdot [1 + j\omega \cdot C_{te}(Z_b(\omega) + Z_e(\omega))], \tag{102}$$

$$Z_b(\omega) = j\omega \cdot l_b + R_s, \tag{103}$$

$$Z_e(\omega) = j\omega \cdot l_e, \tag{104}$$

where l_e and l_b are determined by equations (85) and (86).

$C(\omega_1, \omega_2)$ is the square of the magnitude of the first-order *ac* voltage across the EB junction. As I_C increases, the voltage drop across the EB junction decreases, and thus this term decreases. $L(\omega)$ is proportional to the current gain divided by I_C at the operating frequency, and also decreases as

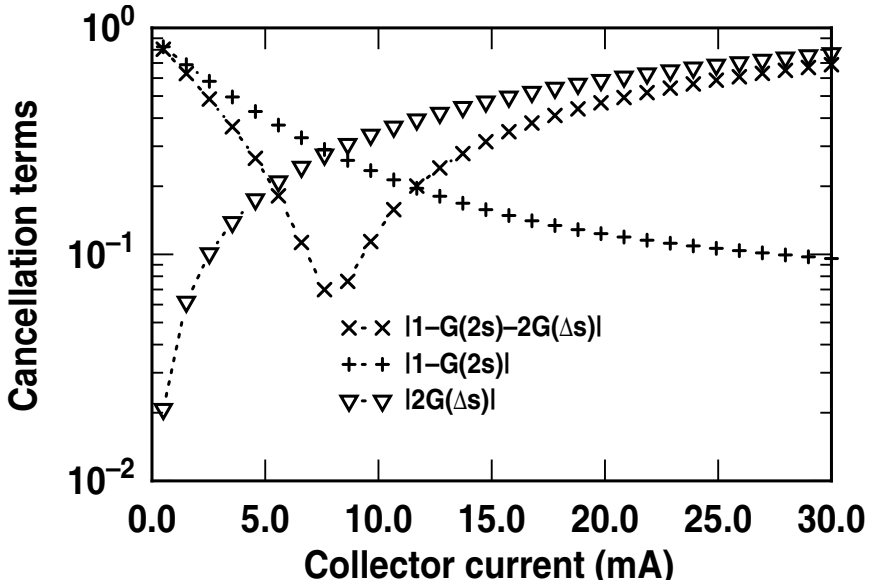

Figure 48: Cancellation term and its two components as a function of I_C. Note that the minimum value of this term responds to the maximum cancellation.

I_C increases. The third term of *IM3*, $|(1 - G(2\omega_1) - 2G(\omega_1 - \omega_2))|$, determines the nonlinearity cancellation, and is called as the cancellation term. Figure 48 shows $|1 - G(2\omega)|$, $|2G(\Delta\omega)|$ and $|1 - G(2\omega) - 2G(\Delta\omega)|$ as functions of I_C. At $I_C =$ 7 mA, $|1 - G(2\omega)|$ is equal to $|2G(\Delta\omega)|$, and they are both in phase. Thus $|1 - G(2\omega) - 2G(\Delta\omega)|$ is minimized at this point.

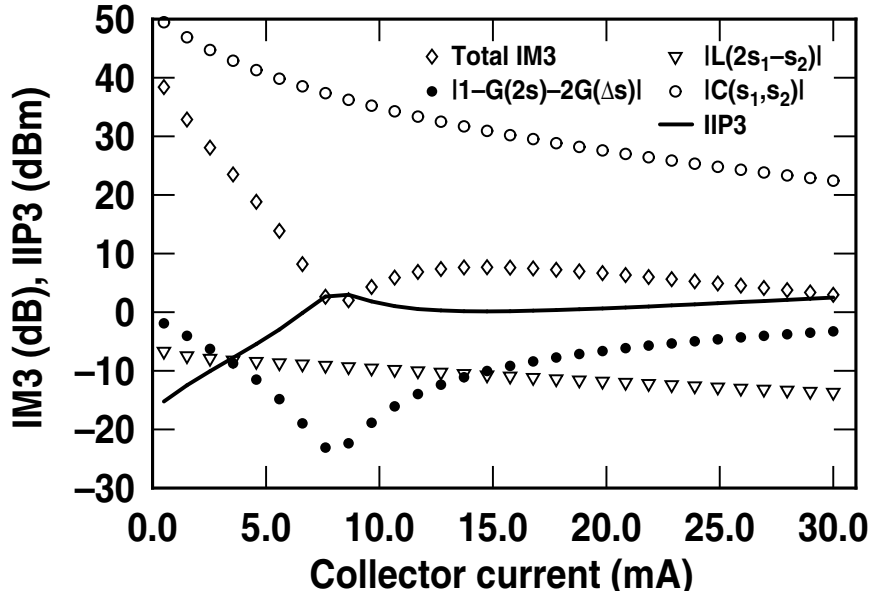

Figure 49: The three terms, the total *IM3*, and the *IIP3* for the input-impedance matched amplifier.

Fixing L_E, the cancellation term can be maximized at a specific I_C. An *IM3* valley or an *IIP3* peak is observed when the cancellation effect dominates. Figure 49 shows the three terms, the total *IM3*, and the *IIP3* for the input-impedance matched amplifier. The minimum value of the cancellation term, *IM3*, and the maximum *IIP3* occur at the same value of I_C, proving that the cancellation term dominates *IM3*.

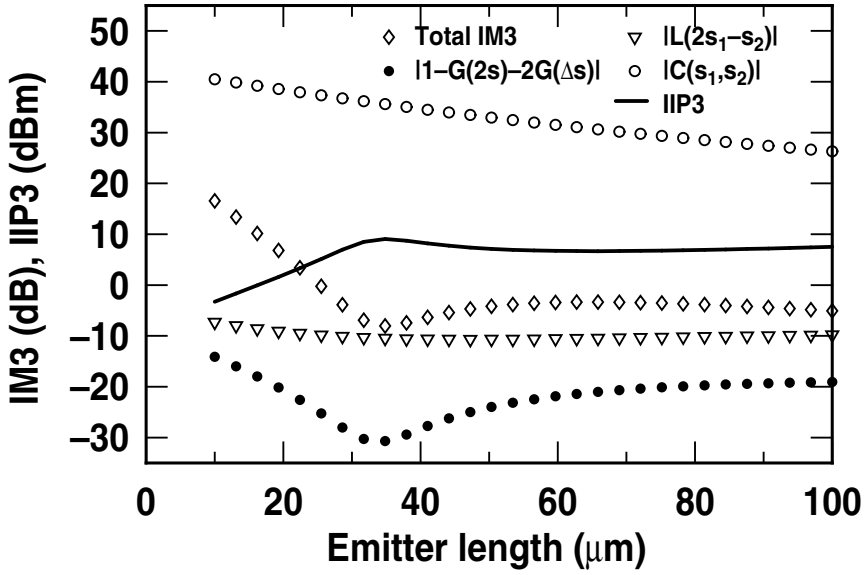

Figure 50: *IM3* terms and the *IIP3* as functions of L_E at $I_C = 6$ mA

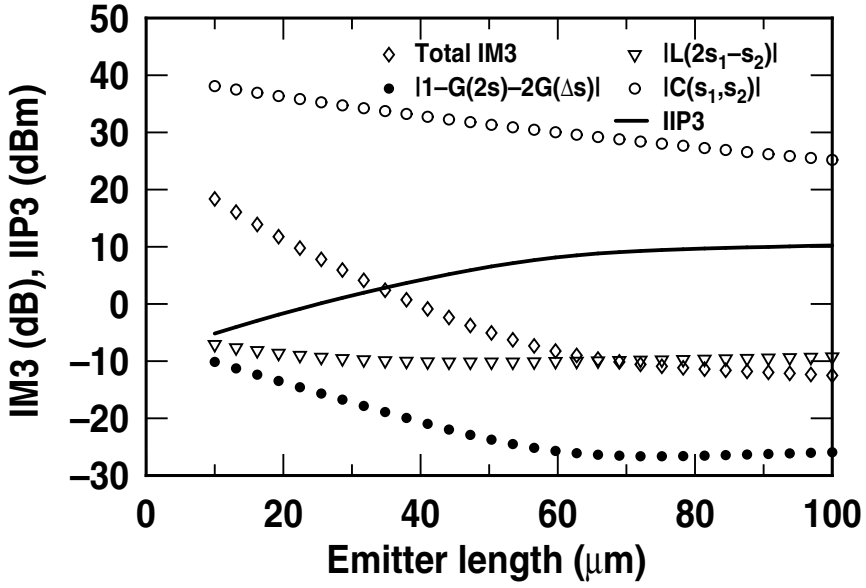

Figure 51: *IM3* terms and the *IIP3* as functions of L_E at $I_C = 7.5$ mA.

The *IM3* is a similar function of L_E. However, at some I_Cs, the other two terms dominate *IM3*. Thus, when I_C is fixed at these values, no *IIP3* peak will be observed. Figures 50 and 51 show the three terms, the *IM3*, and the *IIP3* as functions of L_E at $I_C = 6$ mA and $I_C = 7.5$ mA. An *IIP3* peak can be observed at $I_C = 6$ mA, but no *IIP3* peak is observed at $I_C = 7.5$ mA. The cancellation is pushed to higher current and weakened as I_C increases, hence the other two terms dominate the *IM3* trend.

The derived behavior of gain, *NF*, and *IIP3* for varying L_E and I_C is identical to that obtained from the simulations, although they offer better intuitive insight into the optimum LNA design space. Therefore, one can in the first order locate the optimum design range using the equations above.

CHAPTER V

BEHAVIOR MODEL: A SHORTCUT IN LINEARITY ANALYSIS

5.1 Theory of the lumped-nonlinear-source behavioral model

The key virtue of behavioral model is its simplicity in circuit analysis. However, this advantage is not fully utilized using the linearity analysis approach presented in Chapter III. In a time-domain nonlinear behavioral model of a one-port device, for example, the $I-V$ relation can be characterized as

$$I(t) = f(V(t), V'(t), V''(t), ..., V^{(p)}(t), I'(t), I''(t), ..., I^{(q)}(t)). \tag{105}$$

Using the presented approach (equation (27)), the required node number equals the sum of the orders of the time-derivatives of V and I in function f plus one $(p + q + 1)$, which is comparable with the node number in a physics model. Therefore, the actual simulation time needed using behavioral models may be even longer than the time taken using optimized physics models.

In this chapter, we present a two-port frequency-domain lumped-nonlinear-source (LNS) behavioral model and its associated construction techniques. Using this method, the node number of the system is reduced considerably. Hence, the linearity calculation is simplified.

In a weakly nonlinear system, the equation can be truncated to a polynomial with an order of 3, using the Taylor series. In a periodic-steady-state condition, similar to that used in the Volterra series methodology, the polynomial can be transferred into a Fourier series. The frequencies in the Fourier series includes a base frequency set (e.g., $\omega_1, \omega_2, \omega_3, ...$) and the harmonics (e.g., $2\omega_1, \omega_1 + \omega_2, 3\omega_3 - \omega_2, 0, ...$). As previously, we denote the frequency set that includes all frequencies of the Fourier series as F_A. Then at any frequency $\omega \in F_A$, the response $i(\omega)$ (current in this case) is written as the sum of a linear function of the stimulus at this frequency $v(\omega)$ (voltage in this case), and the second- and third-order intermodulation products:

$$\begin{aligned} i(\omega) &= i_{lin}(\omega) + i_{2nd}(\omega) + i_{3rd}(\omega) \\ &= y_1(\omega)v(\omega) + i_{2nd}(\omega) + i_{3rd}(\omega), \end{aligned} \tag{106}$$

where $y_1(\omega)$ is the admittance, and $i_{2nd}(\omega)$ is the second-order intermodulation at frequency ω. $i_{2nd}(\omega)$ is generated by any two stimuli at frequencies ω_1 and ω_2 which satisfy $\omega_1 + \omega_2 = \omega$, and is written as

$$i_{2nd}(\omega) = \sum_{(\omega_1,\omega_2)\in A_\omega} y_2(\omega_1, \omega_2)v(\omega_1)v(\omega_2), \tag{107}$$

where A_ω is a frequency-pair set: $A_\omega = \{(\omega_1, \omega_2)|\ \omega_1 + \omega_2 = \omega;\ \omega_1, \omega_2 \in F_A\}$. $y_2(\omega_1, \omega_2)$ is the second-order nonlinear parameter, which satisfies $y_2(\omega_1, \omega_2) = y_2(\omega_2, \omega_1)$.

Similarly, $i_{3rd}(\omega)$ is the third-order intermodulation at frequency ω. It is generated by any three stimuli at frequencies ω_1, ω_2, ω_3, which satisfy $\omega_1 + \omega_2 + \omega_3 = \omega$, and is written as

$$i_{3rd}(\omega) = \sum_{(\omega_1,\omega_2,\omega_3)\in B_\omega} y_3(\omega_1, \omega_2, \omega_3)v(\omega_1)v(\omega_2)v(\omega_3), \tag{108}$$

where $B_\omega = \{(\omega_1, \omega_2, \omega_3)|\ \omega_1 + \omega_2 + \omega_3 = \omega;\ \omega_1, \omega_2, \omega_3 \in F_A\}$. $y_3(\omega_1, \omega_2, \omega_3)$ is the third-order nonlinear parameter, which satisfies $y_3(\omega_1, \omega_2, \omega_3) = y_3(\omega_1, \omega_3, \omega_2) = y_3(\omega_3, \omega_1, \omega_2)$.

Equation (106) expresses the presented one-port lumped-nonlinear-source model. In this model, $i_{lin}(\omega) = y_1(\omega)v(\omega)$ is the linear part, while i_{2nd} and i_{3rd} are the lumped nonlinear current sources, which are controlled by stimuli at frequencies in A_ω and B_ω. This model can be extended to N-port nonlinear systems as shown in Figure 30, the only difference being that here the model parameters are frequency dependent. Figure 52 shows a two-port nonlinear system and its equivalent lumped-nonlinear-source model. The stimuli are the voltages, and the responses are the currents of the two ports. In this system, the linear part can be written as $I_{lin} = Y_1 V$, where Y_1 is the Y-matrix,

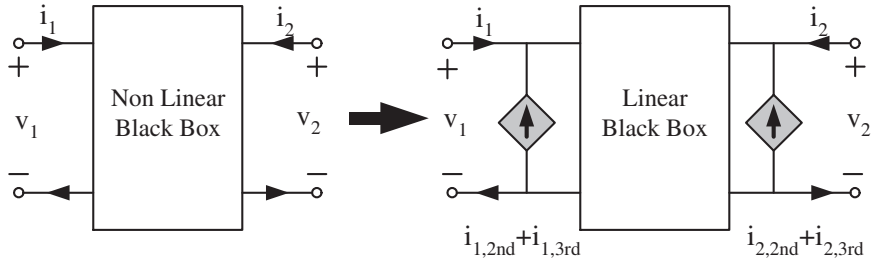

Figure 52: A two-port nonlinear system and its equivalent lumped-nonlinear-source model.

$I_{lin} = [i_{1,lin}, i_{2,lin}]^T$ and $V = [v_1, v_2]^T$. The second- and third-order lumped-nonlinear-sources can

be written as

$$i_{i,2nd}(\omega) = \sum_{(\omega_1,\omega_2)\in A_\omega} \left(\sum_{m,n=1,2} \left(y_{i,2}^{mn}(\omega_1,\omega_2) \times v_m(\omega_1) v_n(\omega_2) \right) \right), \tag{109}$$

$$i_{i,3rd}(\omega) = \sum_{(\omega_1,\omega_2,\omega_3)\in B_\omega} \left(\sum_{m,n,j=1,2} \left(y_{i,3}^{mnj}(\omega_1,\omega_2,\omega_3) \times v_m(\omega_1) v_n(\omega_2) v_j(\omega_3) \right) \right), \; i = 1,2. \tag{110}$$

Observe that $2 \times 2 \times 2 = 8$ parameters are required to model the second-order nonlinear source, and $2 \times 2 \times 2 \times 2 = 16$ parameters are required to model the third-order nonlinear source. For brevity, we denote Y_2 and Y_3, which are $2 \times 2 \times 2$ and $2 \times 2 \times 2 \times 2$ tensors, respectively, as

$$Y_2(i,m,n) = y_{i,2}^{mn}, \tag{111}$$

$$Y_3(i,m,n,j) = y_{i,3}^{mnj}, \; i,m,n,j = 1,2, \tag{112}$$

or

$$Y_2 = \left\{ \begin{bmatrix} y_{1,2}^{11} & y_{1,2}^{12} \\ y_{1,2}^{21} & y_{1,2}^{22} \end{bmatrix} \begin{bmatrix} y_{2,2}^{11} & y_{2,2}^{12} \\ y_{2,2}^{21} & y_{2,2}^{22} \end{bmatrix} \right\}, \tag{113}$$

$$Y_3 = \left\{ \begin{bmatrix} y_{1,3}^{111} & y_{1,3}^{112} \\ y_{1,3}^{121} & y_{1,3}^{122} \end{bmatrix} \begin{bmatrix} y_{1,3}^{211} & y_{1,3}^{212} \\ y_{1,3}^{221} & y_{1,3}^{222} \end{bmatrix} \begin{bmatrix} y_{2,3}^{111} & y_{2,3}^{112} \\ y_{2,3}^{121} & y_{2,3}^{122} \end{bmatrix} \begin{bmatrix} y_{2,3}^{211} & y_{2,3}^{212} \\ y_{2,3}^{221} & y_{2,3}^{222} \end{bmatrix} \right\}. \tag{114}$$

The three parameters $Y_1(\omega_1)$, $Y_2(\omega_1,\omega_2)$, and $Y_3(\omega_1,\omega_2,\omega_3)$ uniquely characterize a nonlinear two-port system. They can be accurately extracted using a set of harmonic balance simulations. The nonlinear current sources can also be transformed into voltage, or hybrid sources. As shown in Figure 53, different forms of two-port lumped-nonlinear-source behavioral models can be used to capture the same device: the Z_1, Z_2, and Z_3 parameters are used to characterize the nonlinear voltage sources of the two ports, while the H_1, H_2, and H_3 parameters are used to characterize the nonlinear current source at port 1 and the nonlinear voltage source at port 2, and the A_1, A_2, and A_3 parameters are used to characterize the nonlinear current and voltage sources at port 2. Each form can be transformed into another with simple mathematics, as derived in Appendix B.

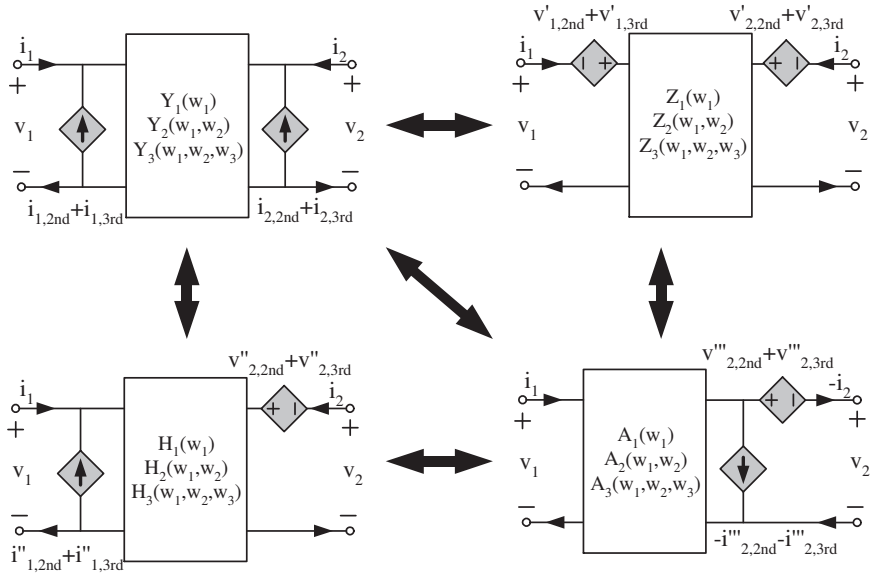

Figure 53: Different forms of two-port lumped-nonlinear-source behavioral models.

Any two parallel, series, or cascade two-port nonlinear systems can be combined into one larger two-port nonlinear system. Figure 54 shows different combinations of two systems using different nonlinear model forms. The Y_1, Y_2, and Y_3 parameters are best suited to a parallel combination; the Z_1, Z_2, and Z_3 parameters are best suited to a series combination; and the A_1, A_2, and A_3 parameters are best suited to a cascade combination. Therefore, more complicated systems can be easily simplified using these transformations and combinations.

5.2 Linearity analysis application

Using this two-port LNS modeling technique, the impact of harmonic source/load impedance on the output *IM2* and *IM3* of a two-port system is investigated. Similar studies on the impact of harmonic impedance on LNAs had been presented in [71]. The goal is to optimize linearity with minimum effect on other system performance metrics such as gain, noise, and input/output match – since the fundamental source/load impedance is fixed. However, those results were limited to a simplified device compact model and circuit topology, and lose accuracy as the equivalent circuit becomes more complicated. Using the presented modeling technique, as we will show, a generic and accurate relation between *IM2* or *IM3* and the source or load harmonic impedances can be obtained regardless of the complexity of the device/circuit. Then a linearity optimization

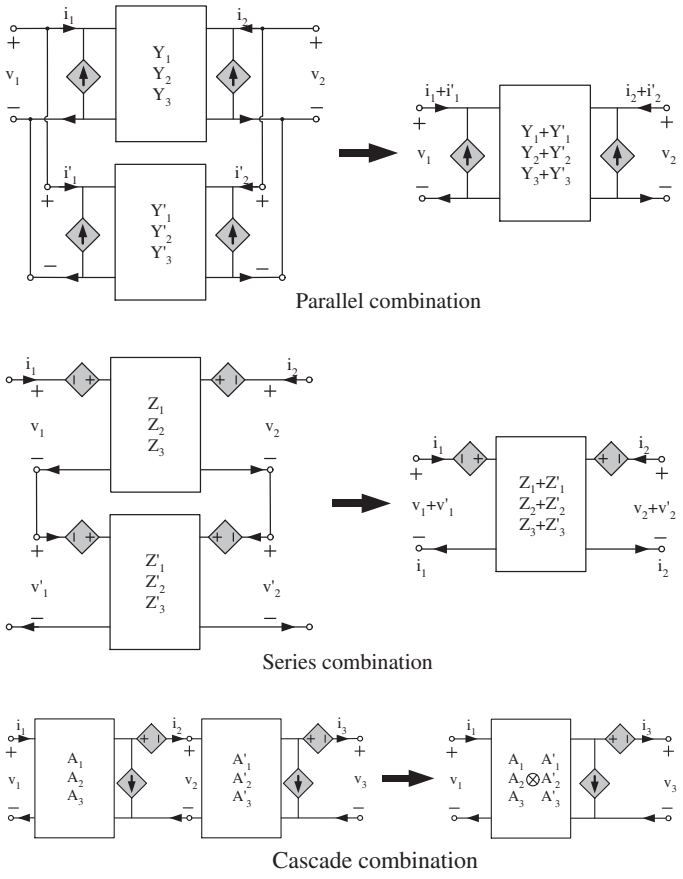

Figure 54: Different combinations of two systems using different nonlinear model forms.

procedure is demonstrated based on this relation. In the last section of this chapter, harmonic load-pull measurements are performed on advanced SiGe HBTs to prove the validity and usefulness of the derived results.

Figure 55 shows the circuit diagram used in the analysis. The input is a two-tone current source in parallel with a source impedance z_S. The output is the power delivered to a load impedance z_L. The two frequencies of the stimulating source (ω_1 and ω_2) are close to each other $|\omega_2 - \omega_1| << \omega_1, \omega_2$. Without loss of generality, we arbitrarily choose a nonlinear two-port system, which can be either a device or a circuit. Its model parameters are Y_1, Y_2, and Y_3.

The *IM2* at $2\omega_1$, and *IM3* at $2\omega_1 - \omega_2$ can be derived using the Volterra-series methodology [36]. In a weakly nonlinear system analysis, one can prove that the impact of source and load impedances at the third order or above harmonic frequencies are negligible. Therefore, only *dc* (or IF) and second-harmonic impedances are considered as the variables that change *IM2* and *IM3*.

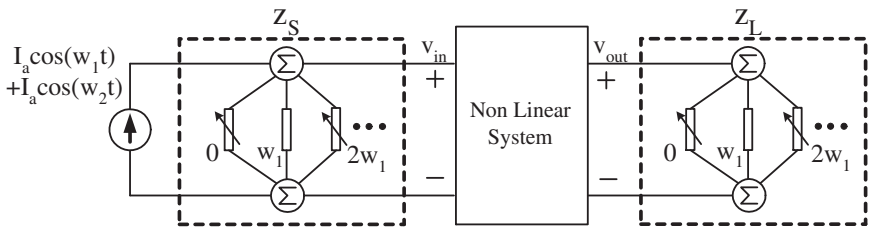

Figure 55: Circuit diagram used in linearity analysis to assess the impacts of harmonic source/load impedances.

First, consider the interferences of the second-harmonic impedance. When the source impedance at each frequency is fixed, the *IM2* can be written as a function of the second-harmonic load admittance $y_{L,2} = y_L(2\omega_1)$

$$IM2(y_{L,2}) \quad = \quad A_{L2}\frac{Real(y_{L,2})}{|y_{L,2} + C_L|^2}, \tag{115}$$

where A_{L2} and C_L are parameters that are derived using the lumped nonlinear current source model (see Appendix B). They are determined by the source impedance and Y_1, Y_2, and Y_3 parameters, but they are not functions of the second-harmonic load impedance $y_{L,2}$. Therefore, the *IM2* as a function of $y_{L,2}$ can be simply expressed as

$$f(x) = A\frac{Real(x)}{|x + C|^2}. \tag{116}$$

Thus, only two variables are required to characterize the *IM2* in the second-harmonic load impedance plane.

IM3 can be written as

$$IM3(y_{L,2}) \quad = \quad A_{L3}\frac{|y_{L,2} + B_{L3}|^2}{|y_{L,2} + C_L|^2}. \tag{117}$$

$$IM2(y_{S,2}) \quad = \quad A_{S2}\frac{|y_{S,2} + B_{S2}|^2}{|y_{S,2} + C_S|^2}. \tag{118}$$

$$IM3(y_{S,2}) \quad = \quad A_{S3}\frac{|y_{S,2} + B_{S3}|^2}{|y_{S,2} + C_S|^2}. \tag{119}$$

where A_{L3} and B_{L3} are parameters that are derived using the lumped nonlinear current source model (see Appendix B). Once again, they are determined by the source impedance and Y_1, Y_2, and

Y_3 parameters, but they are not functions of the second-harmonic load impedance $y_{L,2}$. Therefore, the *IM3* as a function of $y_{L,2}$ can be simply expressed as

$$f(x) = A\frac{|x+B|^2}{|x+C|^2}. \tag{120}$$

Thus, only three variables are required to characterize the *IM3* in the second-harmonic load impedance plane.

Similarly, when the load impedance at each frequency is fixed, both the *IM2* and *IM3* as functions of the second-harmonic source admittance can be simply expressed as in equation (120). In this case, $x = y_{S,2} = y_S(2\omega_1)$ is the second-order source impedance. We write A_{S2}, B_{S2}, and C_S as the A, B, and C parameters of the *IM2* in the second-harmonic source impedance plane; and A_{S3}, B_{S3}, and C_S as the A, B, and C parameters of the *IM3* in the second-harmonic source impedance plane, respectively. These parameters are determined by the load impedance and Y_1, Y_2, and Y_3 parameters.

The impact of the *dc*/IF impedance can be analyzed in the same manner. Note that the *IM2* at $2\omega_1$ is not a function of the IF load/source impedance. However, the *IM3* as a function of the IF load/source impedance can be expressed as in equation (120). We denote $A_{L3,\Delta}$, $B_{L3,\Delta}$, and $C_{L,\Delta}$ as the A, B, and C parameters of the *IM3* in the IF load impedance plane; and $A_{S3,\Delta}$, $B_{S3,\Delta}$, and $C_{S,\Delta}$ as the A, B, and C parameters of the *IM3* in the IF source impedance plane.

Observe that only three parameters are required to characterize the relation between the linearity and harmonic impedance. This expression greatly simplifies the linearity optimization process using harmonic impedance techniques. Moreover, it is universal for all two-port weakly nonlinear systems, and therefore it can be applied in device, circuit, or system linearity optimization methodologies.

5.3 Analysis and simulation

5.3.1 Device analysis

From equations (116) and (120), one can prove that the contours of *IM2* and *IM3* **should be circles or lines** in the load or source harmonic admittance plane (or impedance plane). This is because both the denominators and numerators in the equations are polynomials with orders less than 3.

In the second-harmonic load-pull, $real(C_L)$>0 if the circuit is unconditionally stable. Then, there is **only one peak** ($y_{L,2}$=$conj(C_L)$) and **no valley** of *IM2* in the $y_{L,2}$ plane. The peak (conjugate matching point) can be extracted using the Y_1 parameters and source impedance at $2\omega_1$. In *IM3*, **one peak** ($y_{L,2} \approx -j \times imag(C_L)$) and **one valley** ($y_{L,2} = -B_{L3}$, when $real(B_{L3}) \leq 0$; or $y_{L,2} \approx -j \times imag(B_{L3})$, when $real(B_{L3}) > 0$) exist in the $y_{L,2}$ plane. The same results are observed in the second-harmonic source-pull, except that there is one *IM2* valley in the $y_{S,2}$ plane ($y_{S,2}$=$-B_{S3}$, when $real(B_{S3})\leq 0$, or $y_{S,2}\approx -j\times imag(B_{S3})$, when $real(B_{S3})>0$).

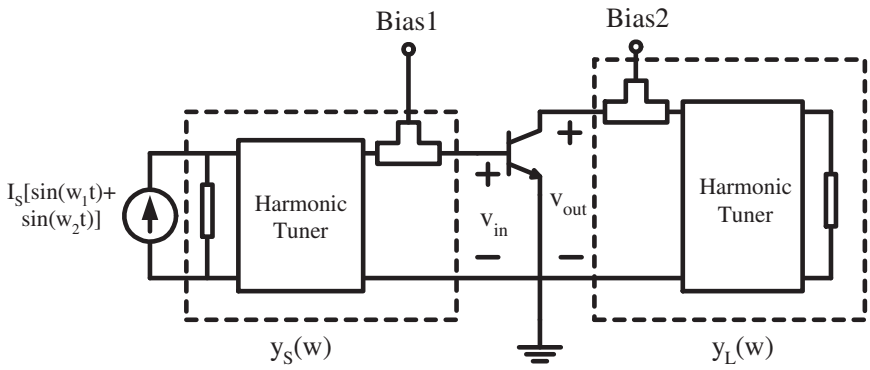

Figure 56: The SiGe HBT used in harmonic-impedance-controlled linearity analysis.

To prove the validity of the conclusions drawn above, we used two-tone harmonic load-pull simulations provided in the HP-ADS tools (harmonic balance approach). The simulation circuit structure is shown in Fig 56. The two fundamental frequencies f_1 and f_2 were set to 1.9 GHz and 1.925 GHz. The sweep range of the second-harmonic load impedance (denoted by $z_{L,2}$) covers the whole Smith chart. *IM2* and *IM3* were then simulated at different harmonic impedances.

Figure 57 shows the simulated *IM2* contours in the $z_{L,2}$ plane. As expected, one maximum point is located in the impedance plane. Observe that these contours are similar to the power match contours. Since the second-order nonlinear sources ($i_{i,2nd}(2\omega_1), i = 1, 2$) are not affected by the harmonic impedances under weakly nonlinear conditions, the second-harmonic power as a function of $z_{L,S}$ is equivalent to conventional load-pull gain matching. Thus, the maximum harmonic gain or *IM2* is achieved when the load impedance conjugate matches the output impedance: $1/C_L$. To minimize *IM2*, as predicted by our theory, the $z_{L,2}$ should be tuned far away from the matching point.

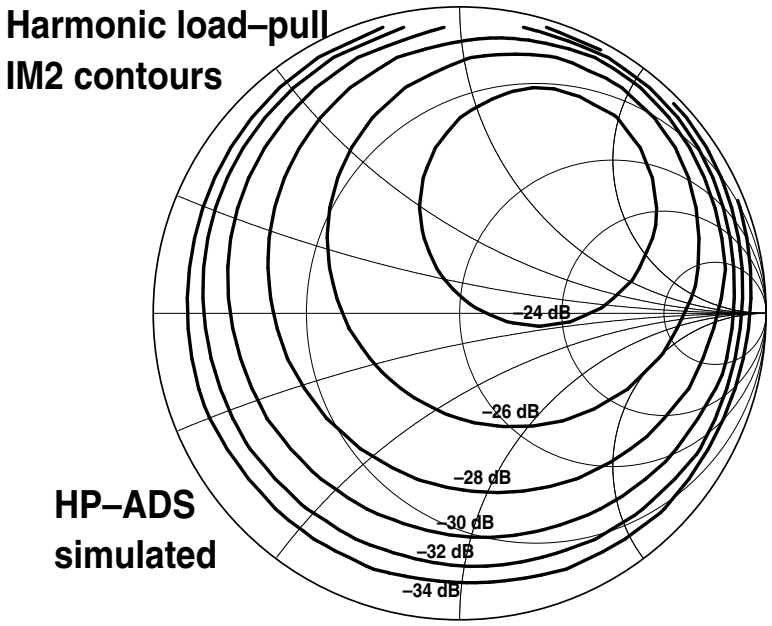

Figure 57: Simulated *IM2* contours in the second-harmonic load impedance plane.

Figure 58 shows the simulated *IM3* contours in the second-harmonic load impedance plane. The results also agree well with the theory: there is one minimum *IM3* point in the $z_{L,2}$ plane, and the maximum point *IM3* is at the edge of the Smith chart (purely inductive or capacitive).

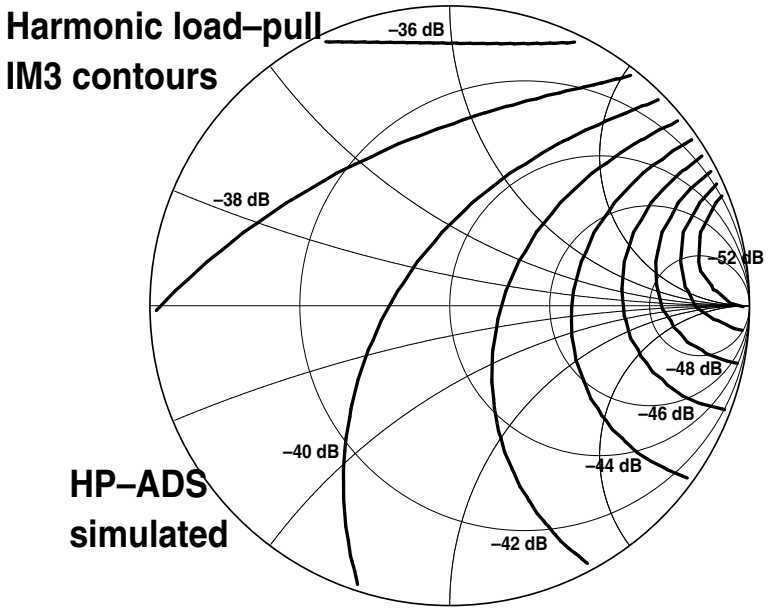

Figure 58: Simulated *IM3* contours in the second-harmonic load impedance plane.

Figures 59 and 60 show the simulated *IM2* and *IM3* contours in the second-harmonic source impedance plane. The patterns are similar to that for *IM3* as a function of load impedance: the peak and valley are both at the edge of the Smith chart. From the plotted contours, one can easily read the optimum $z_{L,2}$ or $z_{S,2}$ to minimize *IM2* and *IM3*, with little effect on the gain and input/output match at the fundamental frequency.

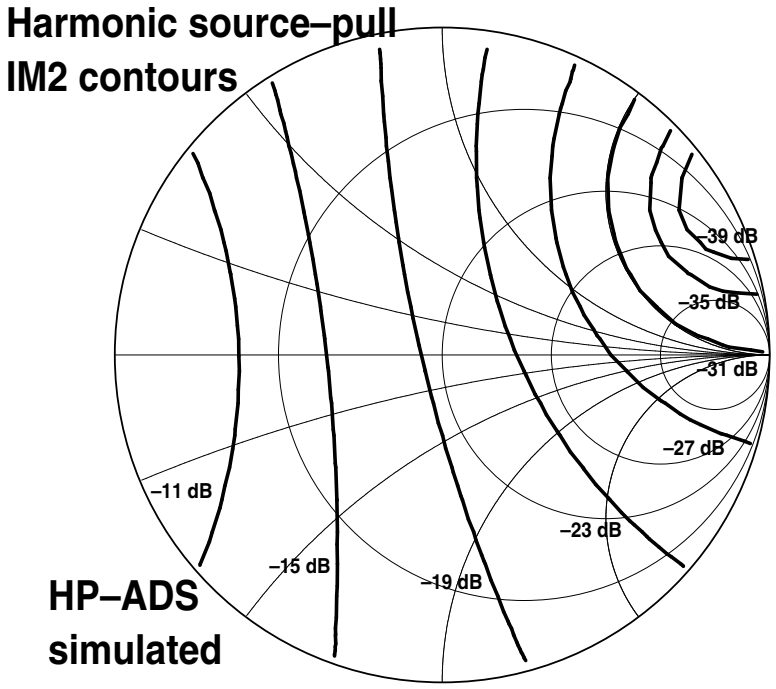

Figure 59: Simulated *IM2* contours in the second-harmonic source impedance plane.

5.3.2 Circuit analysis

The presented methodology is now applied to a simple two-stage amplifier. Figure 61 shows the circuit structure and equivalent cascade two-port systems. The component devices are the same SiGe HBTs used in the analysis above. The *IM2* and *IM3* of the circuit are first simulated using the HP-ADS harmonic balance method. Figures 62 and 63 show the simulated *IM2* and *IM3* contours on the second-harmonic load impedance plane. The results prove that equations (116) and (120) hold in circuit-level analysis.

As a comparison, the *IM2* and *IM3* are also analytically calculated using the transformation and combination theory of the behavioral model. The device behavioral model was extracted using

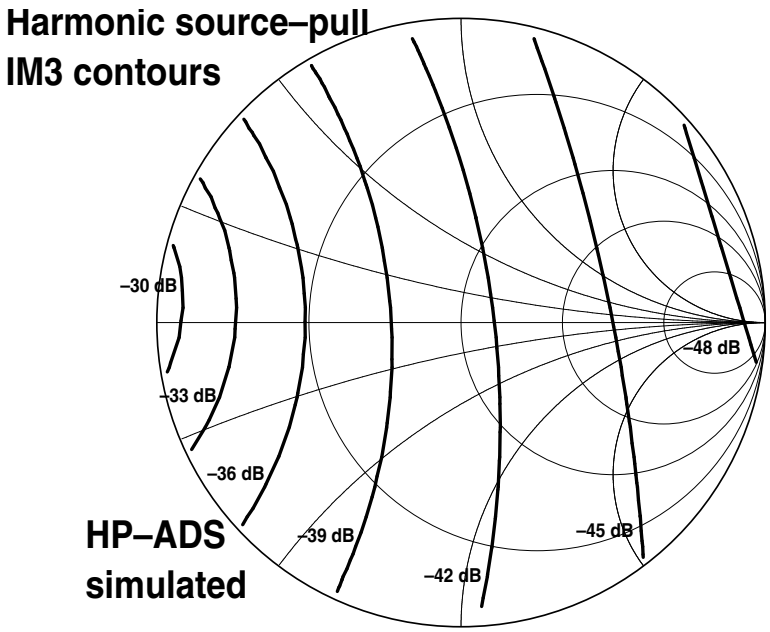

Figure 60: Simulated *IM3* contours in the second-harmonic source impedance plane.

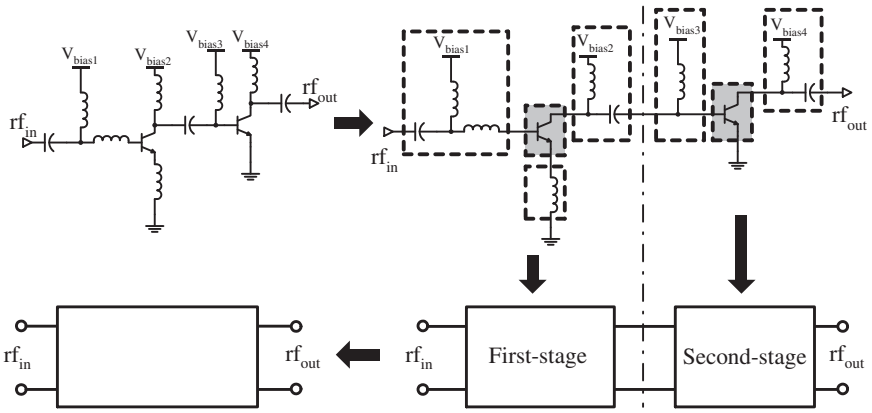

Figure 61: A two-stage amplifier used in harmonic-impedance-controlled linearity analysis.

the previous simulations, then the circuit behavioral model was constructed using parallel, series, and cascade combinations, as illustrated in Figure 61. This approach is much simpler and faster than harmonic balance simulations. Moreover, it can decouple nonlinear sources contributed by individual devices. Note that in this circuit the only two nonlinear contributors are the two SiGe HBTs in the first and second stages. We then analyze the nonlinear contribution, by turning on/off the nonlinear sources of the behavioral model of each stage.

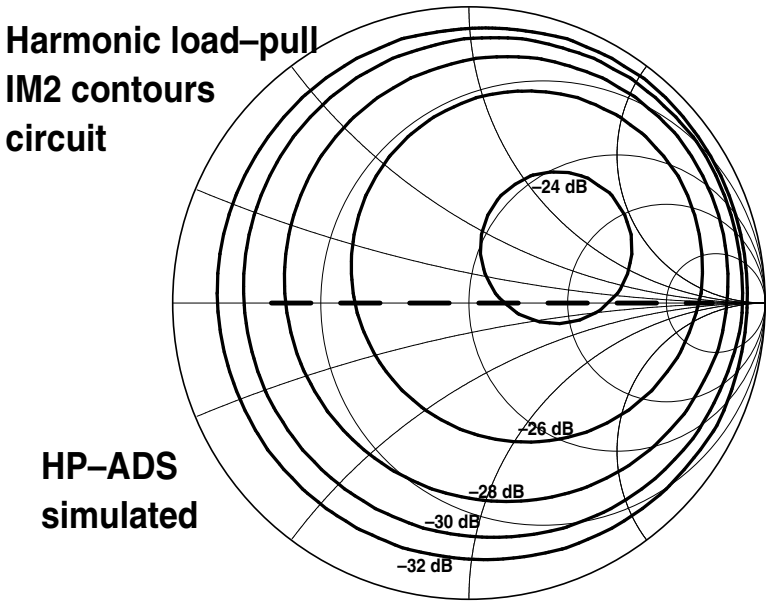

Figure 62: Simulated *IM2* contours of the circuit in the second-harmonic source impedance plane.

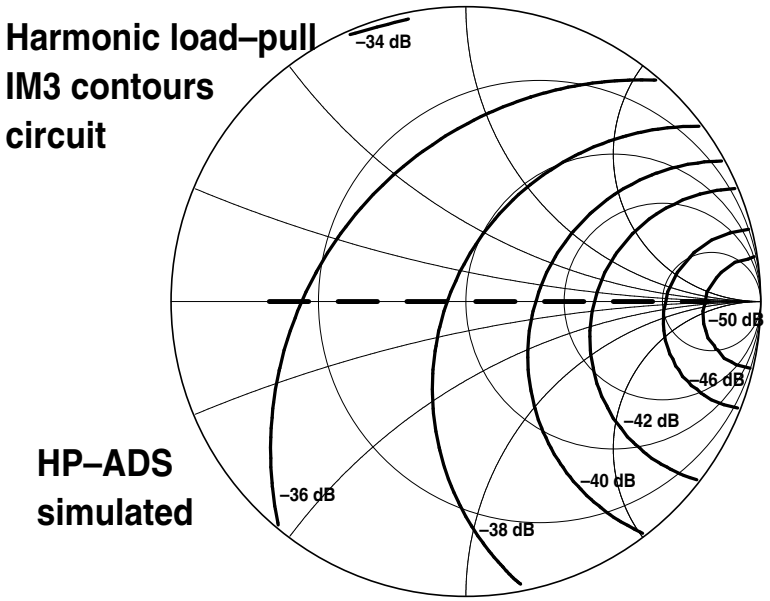

Figure 63: Simulated *IM3* contours of the circuit in the second-harmonic source impedance plane.

Figures 64 and 65 show the *IM2* and *IM3* as functions of the second-harmonic load resistance, calculated using harmonic balance simulation (HP-ADS) and the analytical behavioral model method. The results calculated using the behavioral model method agree well with the results simulated using the harmonic balance method, proving that the presented analytical method is accurate.

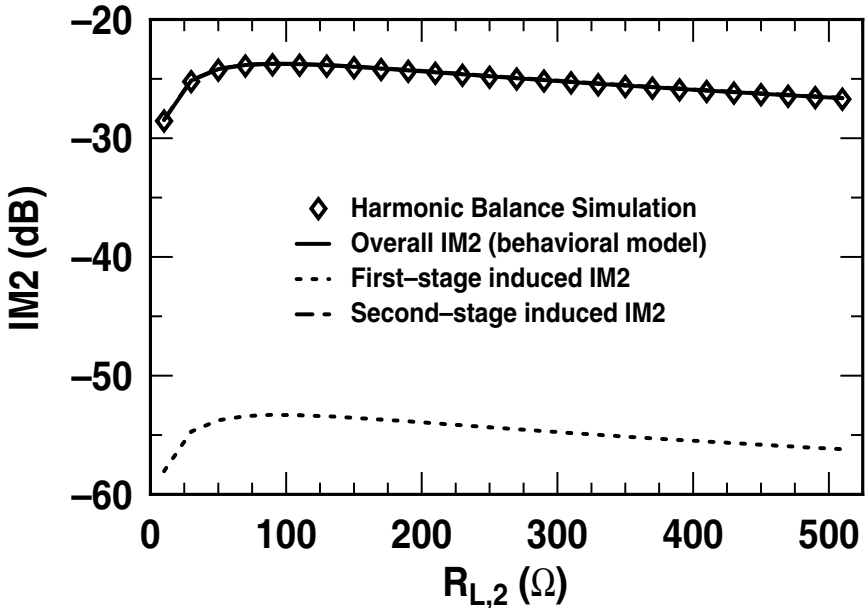

Figure 64: *IM2* as a function of the second-harmonic load resistance, using harmonic balance simulation (HP-ADS), and the behavioral model method.

The results also show that the nonlinearity of the second-stage dominates *IM2* and *IM3*, which is expected since in the cascade amplifier, the input power level at the second-stage is higher than the input power level at the first-stage. Therefore, the second-stage is more vulnerable to device nonlinearity, which explains the similarity between the *IM2*/*IM3* patterns of the circuit and the device (Figure 57 vs. Figure 62, and Figure 58 vs. Figure 63).

Observe that when $R_{L,2} > 200\,\Omega$, the overall *IM3* is much higher than the second-stage *IM3* (by more than 3 dB). This is in contradiction to the conventional cascade theory of the *IIP3* [22], and is caused by the cross-intermodulation term between the first-stage and the second-stage (see the Appendix B for the derivation of this expression). This term induces additional third-order distortion and should be considered in the *IM3* of cascade structures.

5.4 Measurement results

To provide more solid proof to support our theory, harmonic load-pull measurements on devices were also performed. The device used for the measurements was a state-of-the-art SiGe HBT with a peak f_T of 110 GHz. Figure 66 shows the measured f_T and f_{max} as function of J_C of this SiGe HBT technology.

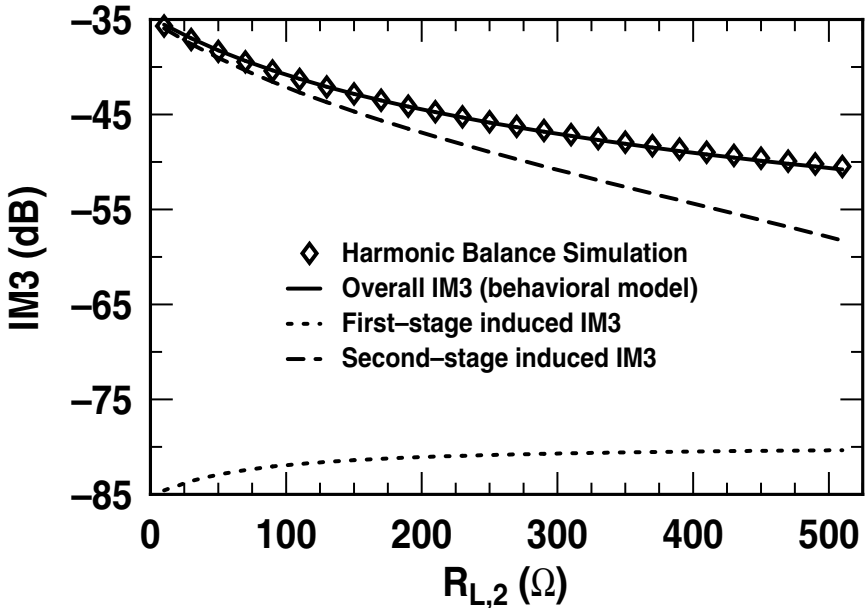

Figure 65: *IM3* as a function of the second-harmonic load resistance, using harmonic balance simulation (HP-ADS), and the behavioral model method.

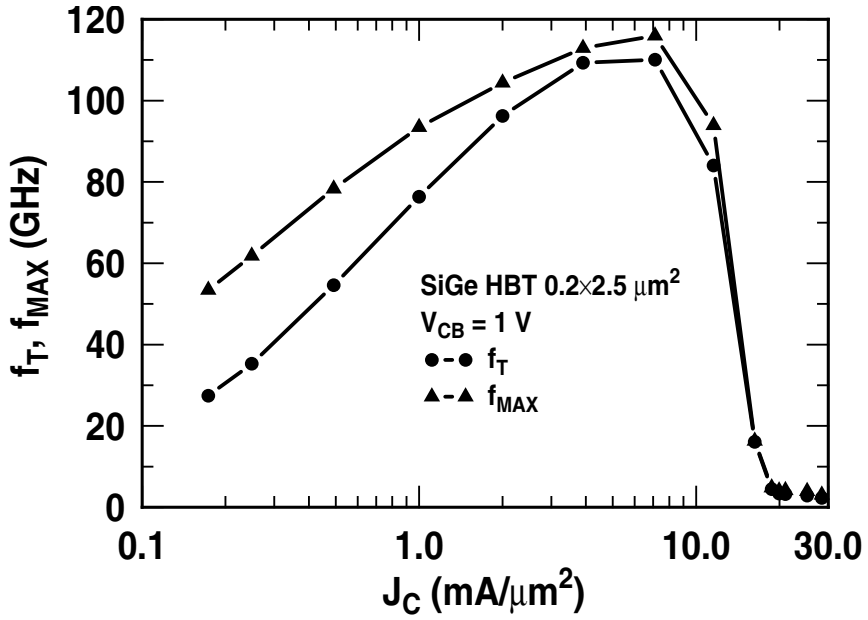

Figure 66: The measured f_T and f_{max} as function of J_C of the SiGe HBT.

The harmonic load-pull was implemented using a cascade tuning technique. The concept of this technique is to increase the tuning capability by sweeping both tuners, such that more combinations of the two tuner states can be obtained to fix the fundamental impedance in a small range. Although

the fundamental impedances are close in these state-pairs, their second-harmonic impedance varies. Therefore, by sweeping these state-pairs, the second-harmonic impedance can be changed while the fundamental impedance remains approximately constant. In this technique, the calibrated s-parameters of the tuners at both the fundamental and the second-harmonic frequencies are needed to calculate the desired sets of state-pairs.

Figure 67 shows the block diagram of the harmonic load-pull measurement setup using the cascaded tuners. Maury mechanical tuners (MT982A02) are chosen to obtain a larger tuning range. The fundamental frequency is fixed at 1.9 GHz, and a spectrum analyzer measures the harmonic output power.

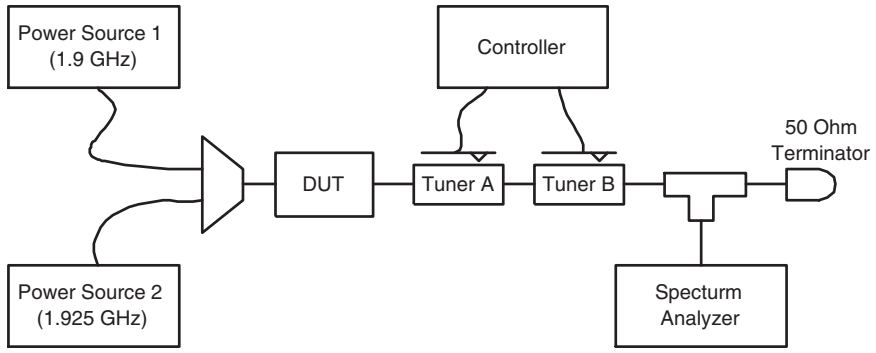

Figure 67: Block diagram of harmonic load-pull measurement setup using cascaded tuners.

Figure 68 shows the measured fundamental and second-harmonic impedances of the chosen set of state-pairs for the harmonic load-pull. The fundamental impedances (∇) are fixed within a small range ($|\Gamma_{L,1} - \Gamma_c| < 0.06$). Note that the sweeping range of the corresponding second-harmonic impedance ($\times$) is limited in a circle ($|\Gamma_{L,2}| < 0.8$). This is caused by the power loss from the cables and probe connected to the load.

Figure 69 shows the measured *IM2* contours in the second-harmonic load impedance plane. As expected, one peak is observed in the impedance plane. Furthermore, to prove the validity of the theory, we extracted C_L in equation (115) using measured *IM2* as a function of harmonic admittance $y_{L,2}$. We then compared the extracted C_L with the measured output impedance of the device at the second-harmonic frequency, which should also be equal to C_L. The results show good agreement: C_L=0.0060+j0.008 from extraction, and C_L=0.0053+j0.0075 from measurement. Therefore, the measured *IM2* contours are consistent with our theory.

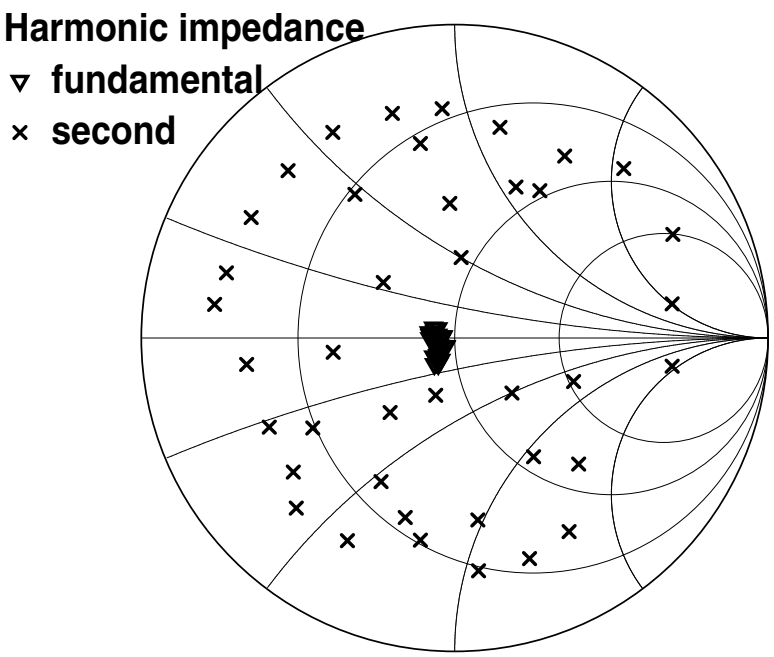

Figure 68: The measured fundamental and second-harmonic impedances used in the harmonic load-pull.

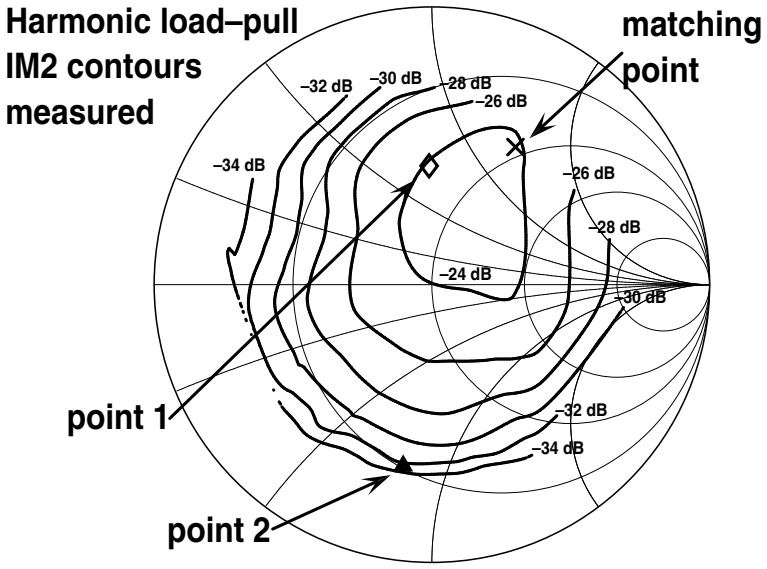

Figure 69: The measured *IM2* contours in the second-harmonic load impedance plane.

Also observe that more than 10 dB improvement of *IM2* can be achieved by tuning the harmonic load impedance. Therefore, the input second-order intermodulation intercept point (*IIP2*) can also be improved by 10 dB. To demonstrate this, two points with different *IM2* were chosen to extract

IIP2.

Figure 70 shows the measured fundamental and second-harmonic output powers as functions of input power at the chosen two points. 10 dB improvement of extracted *IIP2* is obtained when the second-harmonic impedance is set to point 2. Moreover, the second-harmonic power increases about 2 dB as the input power increases 1 dB, meaning that the weakly nonlinear assumption used in the theory is valid at this measurement power level.

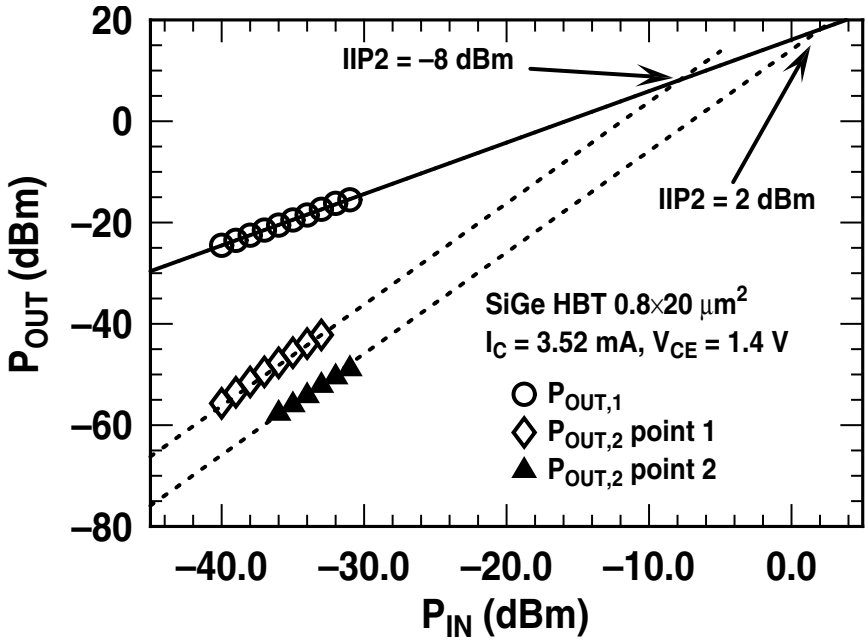

Figure 70: The measured fundamental and second-harmonic output powers as functions of input power at different harmonic impedances.

Figure 71 shows the measured fundamental and third-harmonic output powers as functions of input power at the chosen two points. *IIP3* is improved by 6 dB. Therefore, a better linearity is obtained by changing the harmonic load impedance from point 1 to point 2.

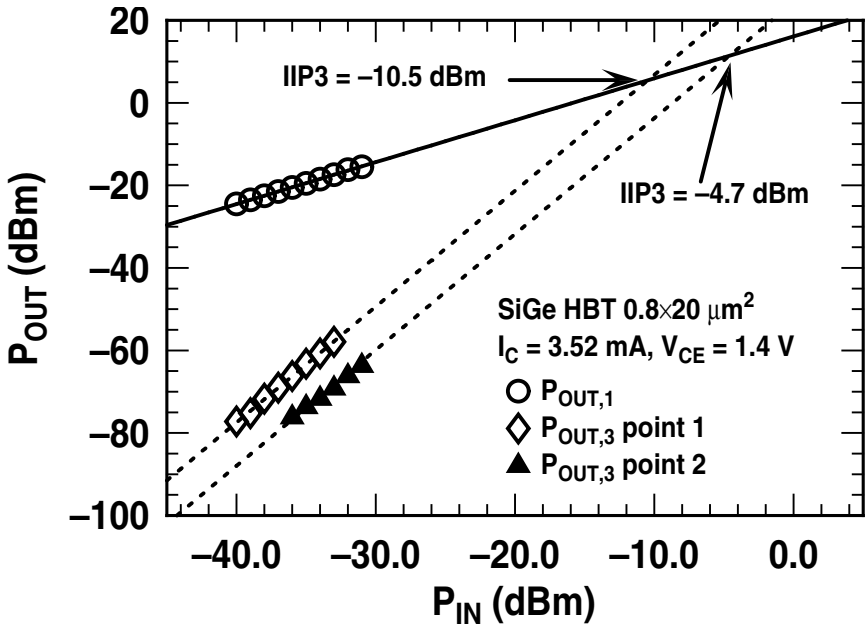

Figure 71: The measured fundamental and third-harmonic output powers as functions of input power at the two points.

CHAPTER VI

RF MEASUREMENT AND CHARACTERIZATION

6.1 On-wafer parasitics and de-embedding techniques

Precise measurement is a prerequisite for accurate device characterization. For accurate evaluation of the high-frequency characteristics and model, s-parameters measurement is a widely-used approach.

As the operating frequency in the measurement increases into the microwave range, the on-wafer parasitics, including the pad-substrate capacitance and wire impedance, become significant. Figure 72 shows a 3-D cross-section view of the on-wafer parasitics and the intrinsic device. As the pad-substrate admittances and wire impedances are comparable to the admittance and impedance of the intrinsic device, respectively, the measured *s*-parameters on the pads (p1 and p2 in Figure 72) misrepresent the device *ac* characteristics. In addition, as the layout device size is normally several hundreds of microns, the distributive nature of the parasitics (coupling between wires and substrate) becomes apparent in the millimeter-wave band.

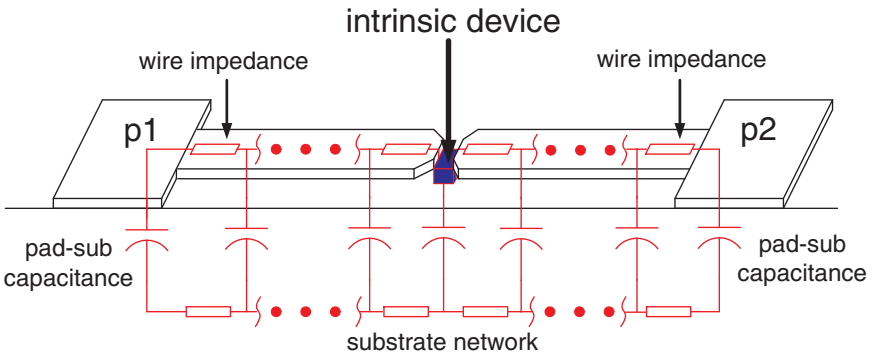

Figure 72: 3-D cross-section view of the on-wafer parasitics and the intrinsic device.

The on-wafer parasitic effects increase as the operating frequency increases. Moreover, the parasitics correlate with the layout and device process (e.g., the properties of the substrate and insulator) and are hard to predict. Therefore, accurate microwave *ac* characterization requires de-embedding

techniques that exclude the parasitic effects and retrieve the intrinsic device characteristics from the measured data.

The standard "open" de-embedding method was first proposed in 1987 [72]. It employs a technique in which the pad-substrate and wire-substrate capacitance are accounted for and calibrated using an OPEN test structure. The layout of the OPEN test structure is the same as the layout of the device under test (DUT), except that the transistor is removed. Figure 73 shows the OPEN test structure and the equivalent circuit used to model the parasitics. Note that in this model, the equivalent circuit of the DUT can be viewed as an intrinsic device in parallel with the OPEN test structure. Then the intrinsic device y-parameters Y^{INT} are derived as [72]

$$Y^{INT} = Y^{DUT} - Y^{OPEN}, \tag{121}$$

where the Y^{DUT} and Y^{OPEN} are the measured y-parameters of the DUT and OPEN, respectively. This approach assumes that the pad-substrate capacitances dominate the parasitics. The validity of this assumption depends on the process technology and layout. Usually, the approach is only accurate at a lower frequency range (i.e., $f < 20$ GHz).

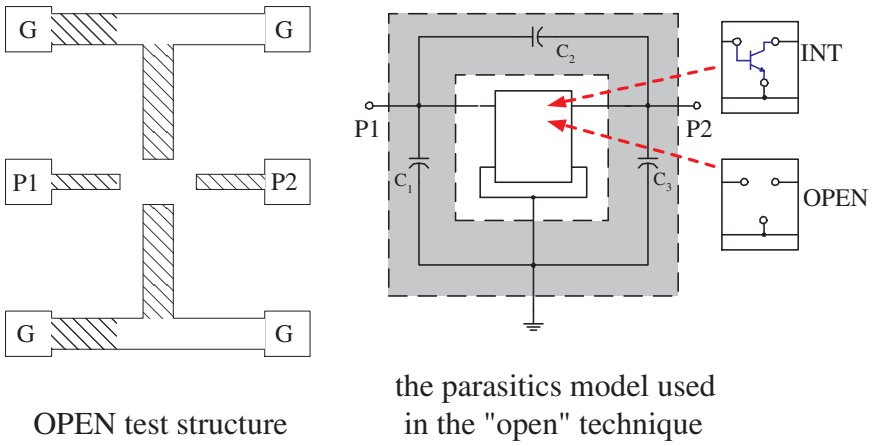

Figure 73: The OPEN test structure and the equivalent circuit of the parasitics model used in the "open" technique.

At higher frequencies (i.e., $f > 20$ GHz), the wire impedance, especially the wire inductance, cannot be neglected. The industry standard "open-short" de-embedding method has thus been introduced to exclude the effects of the wires [25]. Besides the OPEN and DUT, a SHORT test structure is used to extract the parasitics in this method. The layout of the SHORT test structure is

similar to the layout of the OPEN except that the intrinsic ports are both connected to the intrinsic ground. Figure 74 shows the SHORT test structure and the equivalent circuit of the parasitics model. Note that the distributive parasitics are modeled as lumped components in this method. The intrinsic y-parameters can be derived using the measured y-parameters of OPEN (Y^{OPEN}), SHORT (Y^{SHORT}), and DUT (Y^{DUT}) [25]

$$Y^{INT} = \left[\left(Y^{DUT} - Y^{OPEN}\right)^{-1} - \left(Y^{SHORT} - Y^{OPEN}\right)^{-1}\right]^{-1}. \tag{122}$$

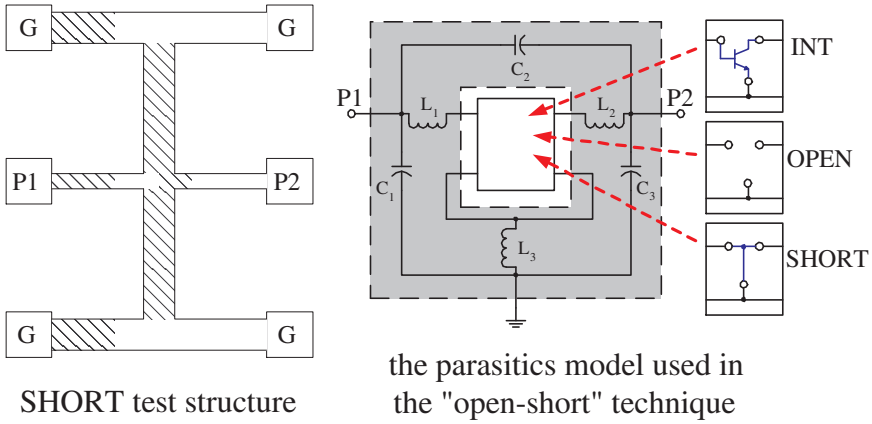

Figure 74: The SHORT test structure and the equivalent circuit of the parasitics model used in the "open-short" technique.

In millimeter-wave circuit applications (e.g., 77 GHz radar in automobile, 60 GHz transceiver with G-bps data-transfer rate), the *ac* characterization at much higher frequencies (i.e., $f > 30$ GHz) is required. However, at this frequency range, the parasitics become more distributive and the "open-short" method starts to lose accuracy. Some high-frequency de-embedding techniques, which use different lumped equivalent circuits to model the parasitics and different test structures for extraction, have been proposed [26][73]. The frequency range of valid *ac* measurement is extended using these methods, but the accuracy of the parasitics model depends on the process technology.

Moreover, some methods that use cascade two-port networks instead of lumped equivalent circuit to model the parasitics are presented [27][74]. Figure 75 shows the parasitics model used in these techniques. The two-port networks capture the distributive effects. However, the model neglects the cross-talk between the two ports, and hence is not suited to devices with lossy substrates.

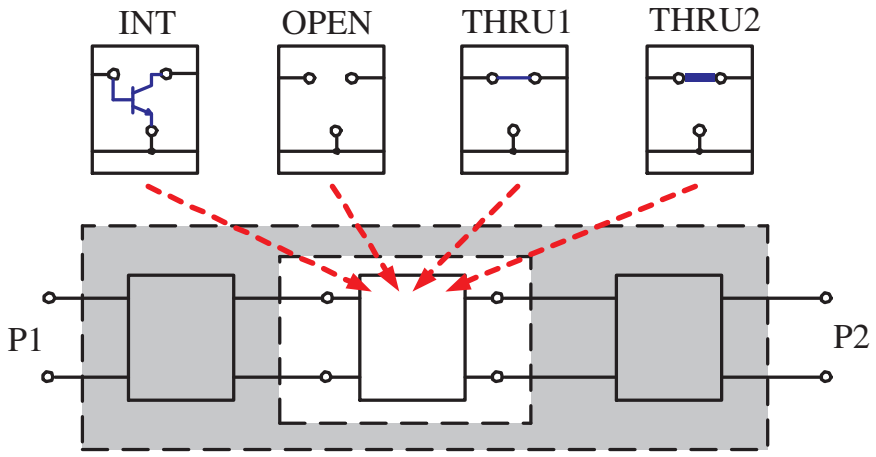

Figure 75: An illustration of the parasitics model using cascade two-port networks.

All the de-embedding techniques introduced above use simplified versions of the parasitics model. These approaches become problematic when the neglected terms in parasitics models are not negligible. To generalize the problem and avoid the potential inaccuracy, a four-port system calibration methodology has been developed [28][75][76]. In this chapter we will focus on the improved four-port de-embedding technique and its application.

6.2 Improved de-embedding technique

6.2.1 Four-port parasitics model and theory

Figure 76 shows a DUT using a four-port network as the parasitics model and a two-port network as the intrinsic device. The four ports include two extrinsic ports (1 and 2), which represent the measured reference-plane, and two intrinsic ports (3 and 4), which represent the intrinsic device characteristics. Since only the I-V characteristics at the extrinsic ports and intrinsic ports are concerned, the four-port network is sufficient to model the parasitics. Furthermore, at small-signal level, the parasitics can be modeled as a linear four-port network, which is characterized by a 4×4 matrix. Then, the I-V relationships of the extrinsic and intrinsic ports can be written as a 4×4 y-matrix according to [75]

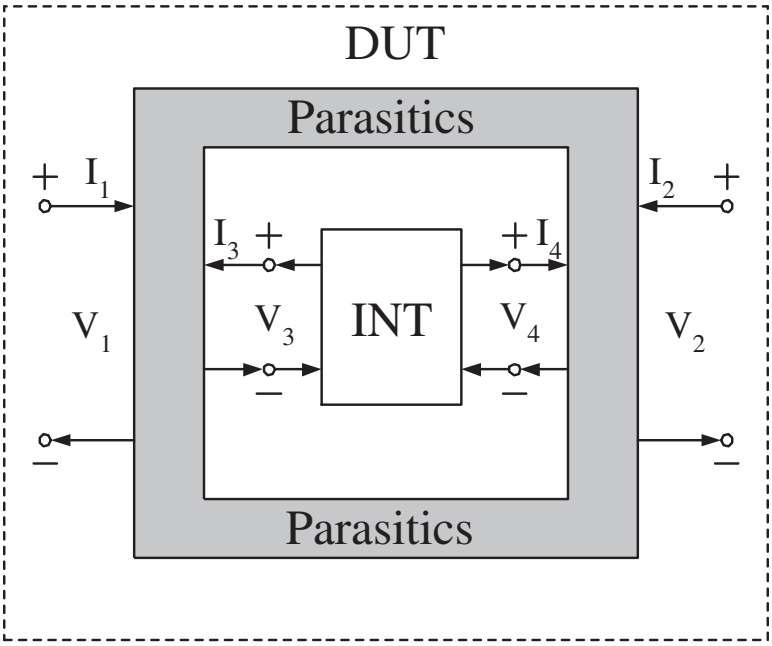

Figure 76: An illustration of the parasitics model using four-port networks.

$$\begin{pmatrix} I_1 \\ I_2 \\ I_3 \\ I_4 \end{pmatrix} = \begin{bmatrix} Y_{11} & Y_{12} & Y_{13} & Y_{14} \\ Y_{21} & Y_{22} & Y_{23} & Y_{24} \\ Y_{31} & Y_{32} & Y_{33} & Y_{34} \\ Y_{41} & Y_{42} & Y_{43} & Y_{44} \end{bmatrix} \begin{pmatrix} V_1 \\ V_2 \\ V_3 \\ V_4 \end{pmatrix}. \tag{123}$$

In some circumstances, Y_{ij} can be ∞ (i.e., there is a short between various ports). In this case, let Y_{ij} be very large to avoid any singularities.

Let V_e and I_e be the extrinsic voltage and current vectors, and V_i and I_i be the intrinsic voltage and current vectors [75]

$$\begin{pmatrix} V_e \\ \\ V_i \end{pmatrix} = \begin{pmatrix} V_1 \\ V_2 \\ V_3 \\ V_4 \end{pmatrix} \text{ and } \begin{pmatrix} I_e \\ \\ I_i \end{pmatrix} = \begin{pmatrix} I_1 \\ I_2 \\ I_3 \\ I_4 \end{pmatrix}. \tag{124}$$

Then we have [77]

$$\begin{pmatrix} I_e \\ I_i \end{pmatrix} = \begin{bmatrix} Y_{ee} & Y_{ei} \\ Y_{ie} & Y_{ii} \end{bmatrix} \begin{pmatrix} V_e \\ V_i \end{pmatrix}, \tag{125}$$

where $[Y_{ee}]$, $[Y_{ei}]$, $[Y_{ie}]$ and $[Y_{ii}]$ are four 2×2 matrices. The extrinsic y-parameters and the intrinsic device y-parameters can then be related as

$$Y^{DUT} V_e = Y_{ee} V_e + Y_{ei} V_i, \tag{126}$$

$$-Y^{INT} V_i = Y_{ie} V_e + Y_{ii} V_i, \tag{127}$$

where Y^{INT} are the intrinsic device y-parameters, and Y^{DUT} are the two-port y-parameters of the DUT.

Note that the current directions of the intrinsic device are opposite to the current directions of the parasitics. It follows from equations (126) and (127) that

$$Y^{DUT} = Y_{ee} - Y_{ei}(Y^{INT} + Y_{ii})^{-1} Y_{ie}, \tag{128}$$

or

$$Y^{INT} = -Y_{ie}(Y^{DUT} - Y_{ee})^{-1} Y_{ei} - Y_{ii}. \tag{129}$$

Once the 16 variables of the 4×4 matrix are known, one can build the appropriate 1-to-1 relationship between the extrinsic and intrinsic y-parameters.

For each test structure, a 2×2 y-parameters can be obtained using the measured raw s-parameters. Given the intrinsic y-parameters of the test structure as well as the measured raw y-parameters, four equations are derived from one test structure. To solve for all 16 variables, one needs to measure at least 4 different test structures unless approximations are made.

For example, the "open-short" de-embedding method uses only two test structures. Appling this method in the four-port system, one gets $V_i = 0$ for the short structure and $I_i = 0$ for the open structure. Applying these two boundary conditions to equation (125) yields

$$Y^{SHORT} = Y_{ee}, \tag{130}$$

$$Y^{OPEN} = Y_{ee} - Y_{ei}(Y_{ii})^{-1} Y_{ie}. \tag{131}$$

Substituting the above equations into equation (128), after simplification, one obtains

$$Y_X + Y_B = Y_X Y_B^{-1} Y_{ie} Y_X^{-1} Y_{ei} + Y_X Y_B^{-1} Y_{ie} Y_B^{-1} Y_{ei}, \tag{132}$$

$$Y_X = Y^{DUT} - Y^{SHORT}, \tag{133}$$

$$Y_B = Y^{SHORT} - Y^{OPEN}. \tag{134}$$

Without loss of generality, Y_X can be any matrix, and thus the equalities above hold when

$$Y_{ie} = Y_{ei} = Y_{ii} = Y^{SHORT} - Y^{OPEN}. \tag{135}$$

Equation (135) gives the condition (assumption) under which the "open-short" approach is valid. At high frequencies (e.g., $f > 30$ GHz), however, this assumption is clearly no longer valid because the distributed nature of the parasitics must be considered.

6.2.2 De-embedding process

As discussed above, at least four test structures are required to solve for all 16 elements. Therefore, besides the OPEN and SHORT, more test structures must be designed to obtain more boundary conditions. The question is then to decide which test structures should be used. Observe in equation (130) that Y_{ee} equals the measured SHORT y-parameters: Y^{SHORT}. To decouple the product term $Y_{ei}(Y^{INT} + Y_{ii})^{-1}Y_{ie}$ in equation (128) is the key to solving Y_{ei}, Y_{ie}, and Y_{ii}. Then, different test structures are chosen to obtain matrices in the form of

$$Y^{LO} = Y_{ei} \begin{bmatrix} 0 & 0 \\ 0 & a \end{bmatrix} Y_{ie}, \tag{136}$$

$$Y^{RO} = Y_{ei} \begin{bmatrix} b & 0 \\ 0 & 0 \end{bmatrix} Y_{ie}, \tag{137}$$

$$Y^{TS} = Y_{ei} \begin{bmatrix} c & c \\ c & c \end{bmatrix} Y_{ie}. \tag{138}$$

where a, b and c are constants. From equation (136)-(138), Y_{ei}, Y_{ie} can be calculated

$$Y_{ie} = k_l Y'_{ei} = k_l \begin{bmatrix} 1 & y_{12}^{RO}/y_{11}^{RO} \\ m_1 & m_1 y_{12}^{LO}/y_{11}^{LO} \end{bmatrix}, \tag{139}$$

$$Y_{ei} = k_r Y'_{ie} = k_r \begin{bmatrix} 1 & m_2 \\ y_{21}^{RO}/y_{11}^{RO} & m_2 y_{21}^{LO}/y_{11}^{LO} \end{bmatrix}. \tag{140}$$

And Y_{ii} is

$$Y_{ii} = k_l k_r Y'_{ie}(Y^{SHORT} - Y^{OPEN})^{-1} Y'_{ei}, \tag{141}$$

where

$$m_1 = \frac{y_{12}^{TS}/y_{11}^{TS} - y_{12}^{RO}/y_{11}^{RO}}{y_{12}^{LO}/y_{11}^{LO} - y_{12}^{TS}/y_{11}^{TS}}, \tag{142}$$

$$m_2 = \frac{y_{21}^{TS}/y_{11}^{TS} - y_{21}^{RO}/y_{11}^{RO}}{y_{21}^{LO}/y_{11}^{LO} - y_{21}^{TS}/y_{11}^{TS}}, \tag{143}$$

and k_l and k_r are scale-factors and will be determined below.

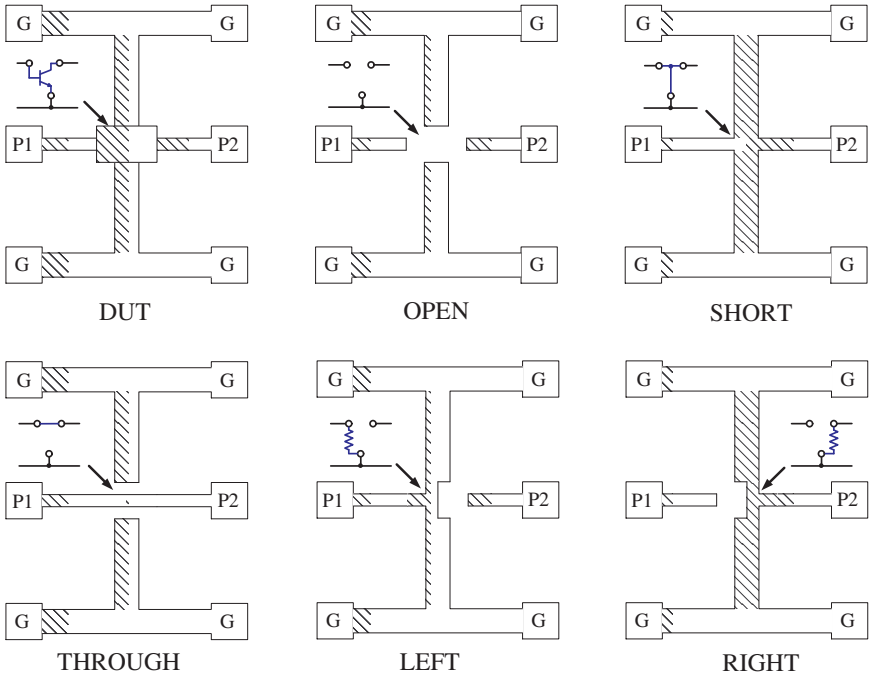

Figure 77: The layout of the DUT and the required test structures used in the four-port technique.

To obtain the matrices discussed above, five test structures are used. Figure 77 shows the layout of the DUT and the required test structures. The $Y_{test,int} + Y_{ii}$ for the OPEN, LEFT, RIGHT, and THROUGH test structures are

$$Y_{open,int} + Y_{ii} = \begin{bmatrix} Y_{ii,11} & Y_{ii,12} \\ Y_{ii,21} & Y_{ii,22} \end{bmatrix}, \tag{144}$$

$$Y_{left,int} + Y_{ii} = \begin{bmatrix} Y_{ii,11} + g_l & Y_{ii,12} \\ Y_{ii,21} & Y_{ii,22} \end{bmatrix}, \tag{145}$$

$$Y_{right,int} + Y_{ii} = \begin{bmatrix} Y_{ii,11} & Y_{ii,12} \\ Y_{ii,21} & Y_{ii,22} + g_r \end{bmatrix}, \tag{146}$$

$$Y_{through,int} + Y_{ii} = \begin{bmatrix} Y_{ii,11} + A & Y_{ii,12} - A \\ Y_{ii,21} - A & Y_{ii,22} + A \end{bmatrix}, A \to \infty, \tag{147}$$

where $g_l = 1/R_l$ and $g_r = 1/R_r$ are conductances of the intrinsic resistors in the LEFT and RIGHT structures, respectively.

Using simple maths, one can prove that [76](see Appendix C for details)

$$Y^{LEFT} - xY^{OPEN} - (1-x)Y^{SHORT} = Y_{ei} \begin{bmatrix} 0 & 0 \\ 0 & a \end{bmatrix} Y_{ie}, \tag{148}$$

$$Y^{RIGHT} - yY^{OPEN} - (1-y)Y^{SHORT} = Y_{ei} \begin{bmatrix} b & 0 \\ 0 & 0 \end{bmatrix} Y_{ie}, \tag{149}$$

$$Y^{THRU} - zY^{OPEN} - (1-z)Y^{SHORT} = Y_{ei} \begin{bmatrix} c & c \\ c & c \end{bmatrix} Y_{ie}, \tag{150}$$

where x, y, and z are the solutions of

$$\left| Y^{LEFT} - xY^{OPEN} - (1-x)Y^{SHORT} \right| = 0, \tag{151}$$

$$\left| Y^{RIGHT} - yY^{OPEN} - (1-y)Y^{SHORT} \right| = 0, \tag{152}$$

$$\left| Y^{THRU} - zY^{OPEN} - (1-z)Y^{SHORT} \right| = 0, \tag{153}$$

and $x \neq 1$, $y \neq 1$, and $z \neq 1$.

Therefore, the normalized Y'_{ei}, Y'_{ie}, and Y'_{ii} are derived. Substituting them into equation (129), one gets

$$Y^{INT} = -k_r k_l Y'_{ie} (Y^{DUT} - Y_{ee})^{-1} Y'_{ei} - Y'_{ii}. \tag{154}$$

The next step is to solve for the scale factors k_r and k_l. Substituting Y^{DUT} using Y^{LEFT} yields

$$Y^{left,int} = -k_r k_l (Y'_{ie} (Y^{DUT} - Y_{ee})^{-1} Y'_{ei} - Y'_{ii}) = k_r k_l Y^{LINT} = \begin{bmatrix} g_l & 0 \\ 0 & 0 \end{bmatrix}, \tag{155}$$

where $Y^{LINT} = -Y'_{ie}(Y^{DUT} - Y_{ee})^{-1} Y'_{ei} - Y'_{ii}$. Thus $k_r k_l = g_l / Y_{11}^{LINT}$. Furthermore, if the parasitics are passive, then $k_r = k_l = \sqrt{g_l / Y_{11}^{LINT}}$.

In summary, the steps in the improved four-port de-embedding process consist of:

1) Measure the s-parameters of the DUT, OPEN, SHORT, THROUGH, LEFT, and RIGHT, and

convert the s-parameters into y-parameters.

2) Solve for x, y, and z in equations (148)−(150), and choose the solution that $x \neq 1$, $y \neq 1$, and $z \neq 1$.

3) Calculate Y^{LO}, Y^{RO}, and Y^{TS} using equations (136)-(138).

4) Obtain normalized $[Y'_{ei}]$, $[Y'_{ie}]$ and $[Y'_{ii}]$ using the equations (139)-(141).

5) Calculate the scale factor $k = k_r k_l$ using equation (155).

6) Calculate the intrinsic y-parameters using equation (129).

6.2.3 Validity and layout concerns

In this method, it has been assumed that equations (144)-(147) and $Y_{ee}+Y_{ei}(Y_{short,int}+Y_{ii})^{-1}Y_{ie} \approx Y_{ee}$ hold, implying that the intrinsic *ac* characteristics of test-structures should be ideal. It is necessary to check the validity of the de-embedding methodology using non-ideal test structures. In ideality, one solution of x, y, and z should be 1, and the intrinsic RIGHT should be open at port 1 and a resistor at port 2. These equations are used for validity verification.

In the current technology for RFIC applications, the intrinsic device layout size is smaller than a few tens of microns, although the DUT size (including pads, etc.) is several hundreds of microns. Thus, the assumptions made in this method are valid at the millimeter-wave band with optimized layout design. In higher frequency measurements (i.e., $f > 300$ GHz), if one can accurately model the non-ideal intrinsic s-parameters of the test structures, the four-port methodology is feasible with a few modifications in the extraction equations.

6.3 Simulation and experiment comparison

Both HP-ADS simulation and *ac* measurement in SiGe HBTs are performed here to fully verify and compare the accuracy of the improved four-port de-embedding methodology with other techniques (e.g., "open-short").

In simulation, several equivalent circuits are chosen to determine how a given parasitic model impacts the four-port technique and "open-short" technique. Figure 78 shows three equivalent circuits of the parasitics. The parasitics models (from 1 to 3) become more distributive, using more capacitors and inductors. The component values in each circuit are extracted and optimized from

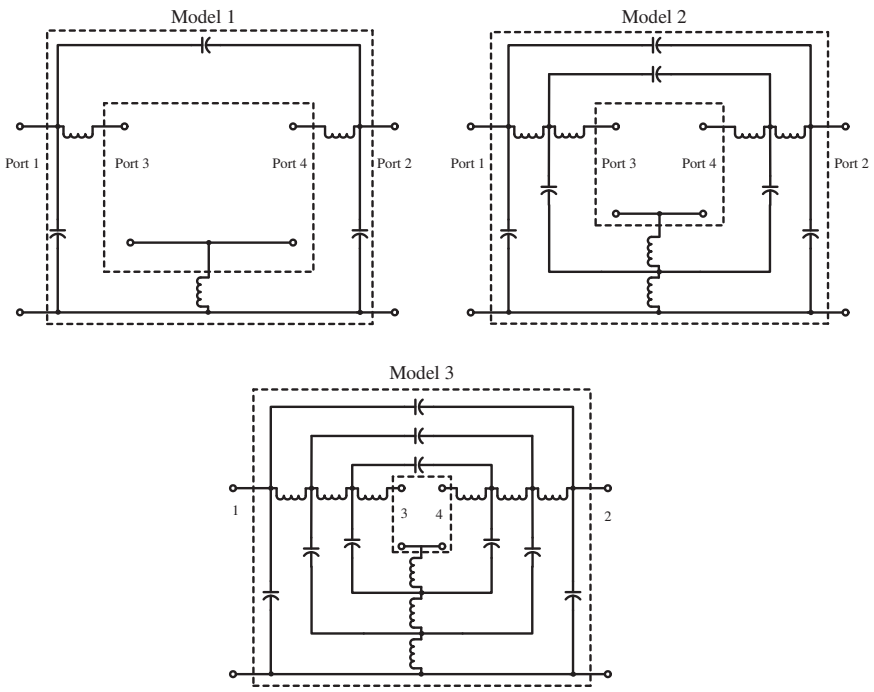

Figure 78: Three equivalent circuits of the parasitics used in the simulation.

the measured s-parameters of parasitics. A device model carefully calibrated to measured data is used to simulate the s-parameters of the SiGe HBTs, both with and without the parasitics. The simulated frequency range is 1 GHz − 100 GHz. Figure 79 shows the de-embedded y-parameters after applying both the "open-short" and the four-port method on each parasitic model.

For equivalent circuit model 1, the intrinsic y-parameters are accurately de-embedded using both the four-port and "open-short" methods. For equivalent circuit models 2 and 3, however, observe that the "open-short" method produces large deviations from the intrinsic y-parameters at frequencies above about 30 GHz. This clearly demonstrates the potential inaccuracy of the traditional "open-short" method at high frequencies. Observe as well that the accuracy of the new four-port method is not dependent on the choice of the equivalent circuit or the frequency.

Both de-embedding techniques applied on actual 2−110 GHz s-parameter measurement data of state-of-the-art SiGe HBTs are also compared. The measured device is a $0.2 \times 2.5\ \mu\text{m}^2$ high-performance *npn* SiGe HBT with a peak f_T of 110 GHz at $J_C = 7.0\ \text{mA}/\mu\text{m}^2$ (the same device used in the harmonic load-pull measurements). The extracted f_T and f_{max} as a function of current density are shown in Figure 66.

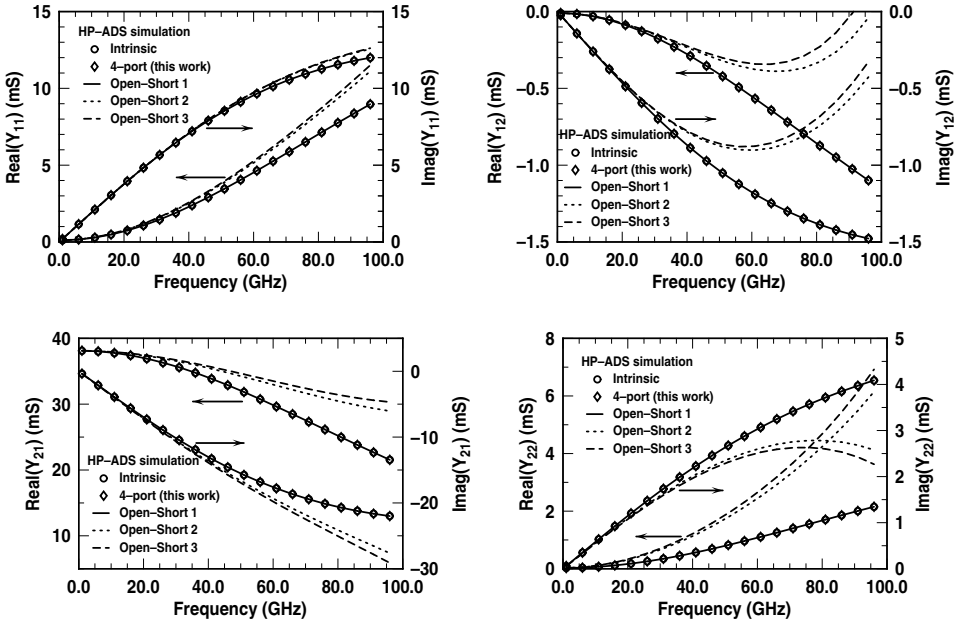

Figure 79: The extracted de-embedded y-parameters using both the "open-short" and the new four-port methods for each parasitic model.

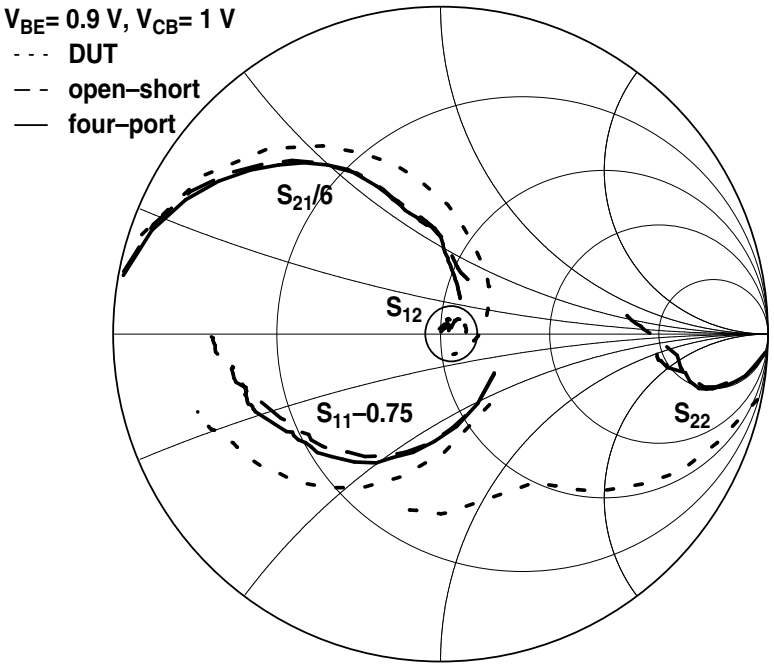

Figure 80: The raw DUT s-parameters and extracted s-parameters using the "open-short" and the four-port methods.

Figure 80 shows the raw DUT s-parameters and extracted s-parameters using the "open-short" and the four-port method. For a better comparison, we have plotted $S_{21}/6$ and $S_{11}-0.75$ rather than

S_{21} and S_{11}. Observe that there are large deviations between the un-deembedded and de-embedded data, indicating that the on-wafer parasitics are significant in this SiGe technology. Note that the de-embedded s-parameters using the two methods are in close agreement, except in the high-frequency range, confirming the validity of both methods at low frequencies.

To more closely examine the differences between the two methods, we plot the de-embedded Y_{21} as a function of frequency, as shown in Figure 81. The deviation of the results is negligible at frequencies lower than about 30 GHz. At frequencies higher than 30 GHz, the "open-short" method overestimates the magnitude and underestimates the phase of Y_{21}. This is caused by the distributive nature of the wire lines between the pads and the intrinsic device. As expected, the error increases as the frequency increases.

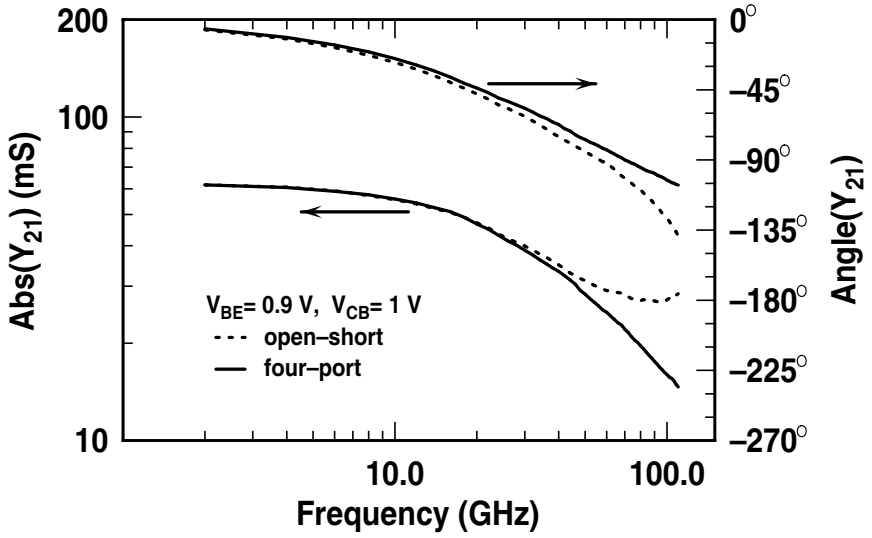

Figure 81: The extracted y_{21} as a function of frequency using the "open-short" and the four-port methods.

These errors can severely distort the measured characteristics of the device (e.g., the gain) at high frequencies. Figure 82 shows the current gain H_{21} as a function of frequency at different bias points. Although the current gains extracted using the two methods nearly overlap at lower frequencies (the "open-short" de-embedded gain is slightly less than the "four-port" gain), at frequencies above 70 GHz, the current gain extracted with the "open-short" method ceases to decrease, which is clearly not congruent with the physical (real) behavior of the intrinsic device. Observe, however, that the current gain with the four-port method continues decreasing with a constant slope of about

20 dB per decade, indicating the accuracy of this technique.

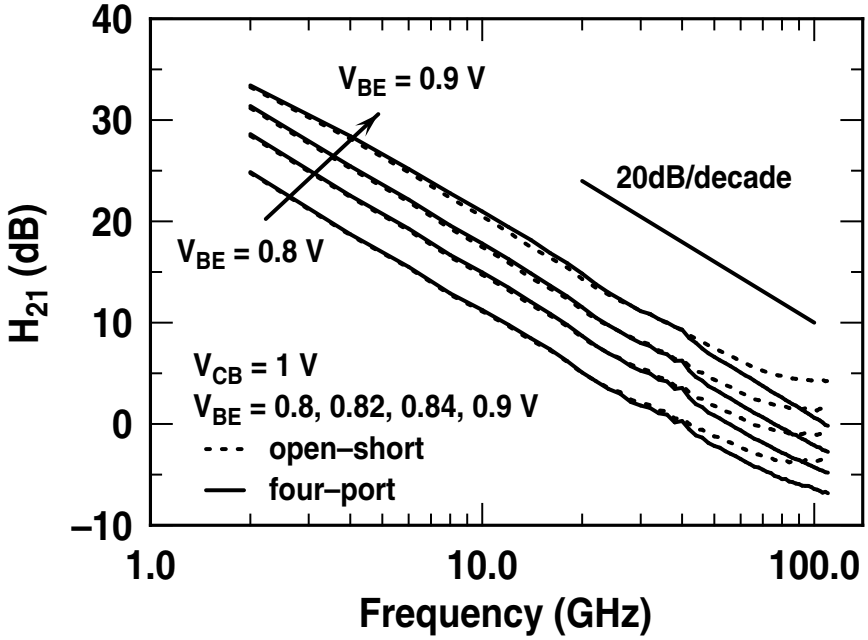

Figure 82: The current gain H_{21} as a function of frequency at different bias points.

6.4 Noise de-embedding technique

The four-port technique can also be applied to noise de-embedding. Recall the n-port noise model illustrated in Chapter III, where a generalized noisy system can be characterized by a noise current correlation matrix SY_n or noise voltage correlation matrix SV_n. The relation between the current correlation matrix and the voltage correlation matrix can be written as

$$SY_n = \begin{bmatrix} \widetilde{i_{n,1}}\widetilde{i^*_{n,1}} & \cdots & \widetilde{i_{n,1}}\widetilde{i^*_{n,n}} \\ \vdots & \ddots & \vdots \\ \widetilde{i_{n,n}}\widetilde{i^*_{n,1}} & \cdots & \widetilde{i_{n,n}}\widetilde{i^*_{n,n}} \end{bmatrix} = Y \times SV_n \times Y^*, \tag{156}$$

where $\widetilde{i_{n,j}},\ j = 1, 2, 3, \ldots$ is the frequency-domain noise current source at port j. Y is the Y-matrix of this n-port system. In a two-port system, the minimum noise figure F_{min}, noise impedance R_n, and the optimum noise admittance Y_{opt} can be directly converted into the noise current correlation matrix SY_2 [61].

Once the 4 × 4 Y-matrix is obtained, the noise de-embedding method is straight-forward [78]. Figure 83 shows the equivalent circuit of the noise model of the four-port parasitics network and

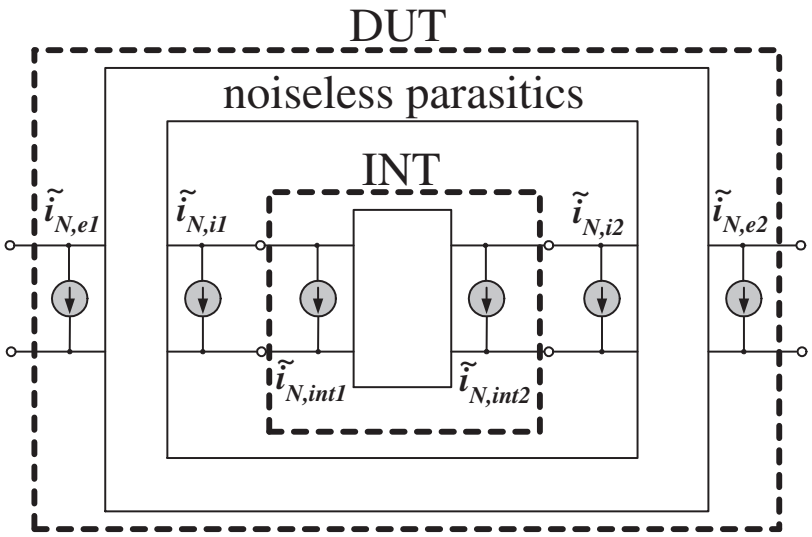

Figure 83: The equivalent circuit of the noise model of the four-port parasitics network and two-port intrinsic device.

two-port intrinsic device. The noise current correlation matrix SY_4 of the parasitics can be written as

$$SY_4 = \begin{bmatrix} SY_{n,11} & SY_{n,12} & SY_{n,13} & SY_{n,14} \\ SY_{n,21} & SY_{n,22} & SY_{n,23} & SY_{n,24} \\ SY_{n,31} & SY_{n,32} & SY_{n,33} & SY_{n,34} \\ SY_{n,41} & SY_{n,42} & SY_{n,43} & SY_{n,44} \end{bmatrix} = \vec{i_n}\vec{i_n^*}, \tag{157}$$

where $SY_{n,ij},\ i,j = 1,2,3,4$ are the noise current correlation between port i and port j. $\vec{i_n}$ is written as

$$\vec{i_n} = \begin{pmatrix} \widetilde{i_{n,1}} \\ \widetilde{i_{n,2}} \\ \widetilde{i_{n,3}} \\ \widetilde{i_{n,4}} \end{pmatrix} = \begin{pmatrix} \widetilde{i_{n,e}} \\ \\ \widetilde{i_{n,i}} \end{pmatrix}, \tag{158}$$

Then, if the four-port system is passive, one obtains

$$SY_4 = \begin{bmatrix} SY_{n,ee} & SY_{n,ei} \\ SY_{n,ie} & SY_{n,ii} \end{bmatrix} = 4kTReal\left(\begin{bmatrix} Y_{ee} & Y_{ei} \\ Y_{ie} & Y_{ii} \end{bmatrix}\right), \tag{159}$$

where $\widetilde{i_{n,e}}$ and $\widetilde{i_{n,i}}$ are extrinsic and intrinsic noise current sources, respectively.

The four-port I-V relation of the DUT, considering noise currents, can then be written as

$$\begin{pmatrix} I_e + \widetilde{i_{n,e}} \\ I_i + \widetilde{i_{n,i}} + \widetilde{i_{n,int}} \end{pmatrix} = \begin{bmatrix} Y_{ee} & Y_{ei} \\ Y_{ie} & Y_{ii} \end{bmatrix} \begin{pmatrix} V_e \\ V_i \end{pmatrix} \tag{160}$$

One can thus calculate the intrinsic noise correlation matrix as (see Appendix C for details)

$$SY_{n,int} = (Y_T)^{-1}(SY_{n,total} - SY_{n,ee})(Y_T^*)^{-1} - SY_{n,ii} + (Y_T)^{-1}SY_{n,ei} + SY_{n,ie}(Y_T^*)^{-1}. \tag{161}$$

where $Y_T = Y_{ei}(Y^{INT} + Y_{ii})^{-1}$.

The noise characteristics were also simulated and compared in HP-ADS on the same device. Figure 84 shows the noise characteristics of a DUT, both intrinsic and de-embedded, using both the traditional "open-short" and the new four-port method. The results again show good precision using the four-port method. For the "open-short" technique, NF_{min} and Γ_{opt} are de-embedded correctly for a wider frequency range (up to 60 GHz). Note, however, that R_n is underestimated at frequencies above about 30 GHz. Hence for an accurate noise characterization of SiGe HBTs at high frequencies, the new four-port de-embedding methodology is again preferred.

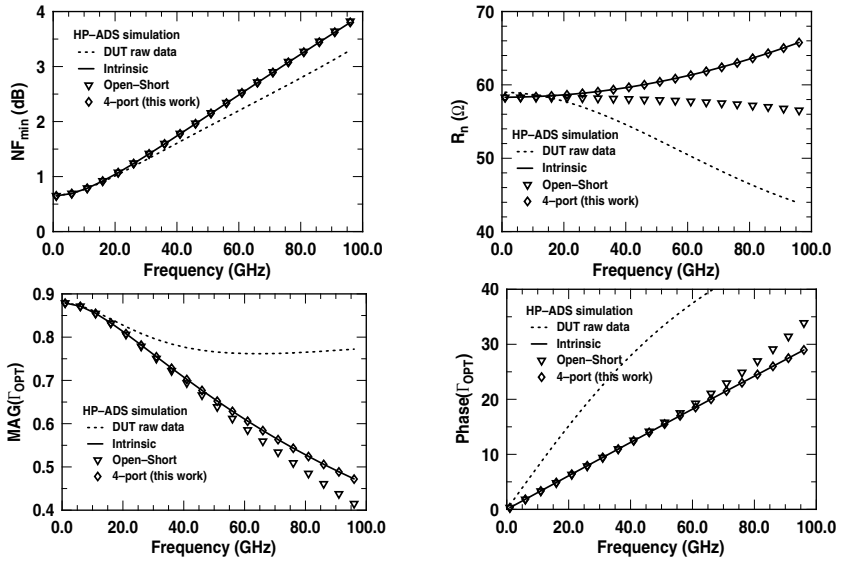

Figure 84: The noise characteristics of the DUT, both intrinsic and de-embedded using the "open-short" method and the proposed four-port method.

6.5 *Passive-substrate-coupling characterization*

The substrate coupling in SiGe HBTs is significant in the millimeter-wave band because of the low resistivity of the Si-bulk. This parasitic severely degrades the isolation between RF building blocks (e.g., LNAs to PAs, mixers to oscillators) at higher frequencies and raises concerns on both circuit design and modeling [79][80]. Therefore it is clearly important to accurately characterize the substrate coupling.

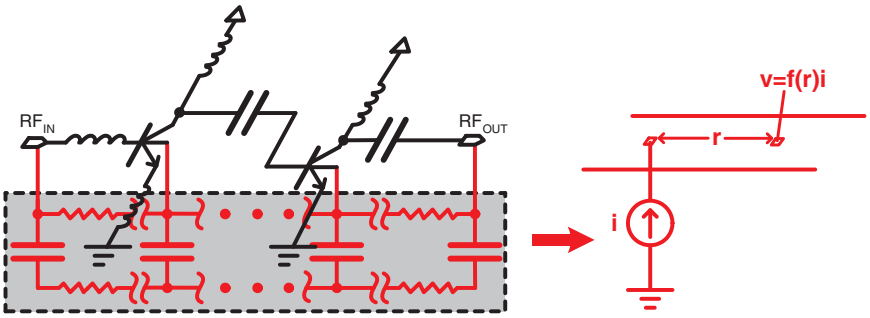

Figure 85: The models used in the substrate coupling. The equivalent circuit method is on the left, and the Green-function method is on the right

Similar to the device models discussed in Chapter III, there are two methods used to model the substrate coupling: the equivalent circuit approach (e.g., resistor-capacitor networks) [11] and the behavioral Green-function approach [80]. Figure 85 shows the two methodologies used to model the substrate coupling.

In the Green-function model, the coupling is characterized by a function $G(\vec{r_1}, \vec{r_2})$: suppose there is a current source applied at the location $\vec{r_1}$, then the responding voltage sensed at the location $\vec{r_2}$ will be $v(\vec{r_2}) = G(\vec{r_1}, \vec{r_2})i(\vec{r_1})$. In general, the Green function G is frequency-dependent to capture the impact of the frequency on loss and phase-delay of the signal. In a homogeneous substrate, this function only depends on the distance between $\vec{r_1}$ and $\vec{r_2}$

$$G(\vec{r_1}, \vec{r_2}) = f(|\vec{r_1} - \vec{r_2}|) = f(r), \tag{162}$$

where $r = |\vec{r_1} - \vec{r_2}|$. In a linear system, one can use superposition to calculate the coupling of multiple interference sources. Therefore, the behavior of the substrate coupling can be modeled as a simple function $f(r)$.

Apparently, the $f(r)$ can be extracted using measured z-parameters (z_{21} or z_{12}). Again, accurately extracting the parameters in the millimeter-wave band is a challenge in RF measurements. As demonstrated previously, the four-port methodology is preferred for high-frequency characterization. However, since now the r can be much larger than the dimension of a device, the previously-presented four-port de-embedding technique may fail because of the the non-idealness of the THROUGH test structure. Here, we present a modified four-port technique that does not require the THROUGH test structure and thus is feasible for the characterization of the substrate coupling.

In this modified approach, we assume that the parasitics are passive. Therefore, the y-parameters of the four-port matrix satisfy $y_{ij} = y_{ji},\ i \neq j$. Then $Y_{ei} = Y_{ie}^T$. Observe in equations (140) and (139) that y_{21}^{RO}, y_{11}^{RO}, y_{21}^{LO}, and y_{11}^{LO} can be extracted using the OPEN, SHORT, LEFT, and RIGHT test structures.[1] The question is how to extract k_r and m_2 (m_1 and k_l should equal to m_2 and k_r in the passive system, respectively) without using the THROUGH test structure.

From equations (128) and (130), we get

$$(Y^{LEFT} - Y^{SHORT})^{-1} - (Y^{OPEN} - Y^{SHORT})^{-1} = Y_{ie}^{-1} Y_{left,int} Y_{ei}^{-1}, \tag{163}$$

$$(Y^{RIGHT} - Y^{SHORT})^{-1} - (Y^{OPEN} - Y^{SHORT})^{-1} = Y_{ie}^{-1} Y_{right,int} Y_{ei}^{-1}. \tag{164}$$

For brevity, we rewrite Y_{ei} and Y_{ie} as

$$Y_{ei} = \begin{bmatrix} c_1 d & e \\ d & c_2 e \end{bmatrix} = Y_{ie}^T. \tag{165}$$

where $c_1 = y_{11}^{RO}/y_{21}^{RO}$, $c_2 = y_{21}^{LO}/y_{11}^{LO}$, $e = k_r m_2$, and $d = k_r/c_1$. And we denote $Z^{LX} = (Y^{LEFT} - Y^{SHORT})^{-1} - (Y^{OPEN} - Y^{SHORT})^{-1}$ and $Z^{RX} = (Y^{RIGHT} - Y^{SHORT})^{-1} - (Y^{OPEN} - Y^{SHORT})^{-1}$. Recall that

$$Y_{open,int} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}, \tag{166}$$

$$Y_{left,int} = \begin{bmatrix} g_l & 0 \\ 0 & 0 \end{bmatrix}, \tag{167}$$

[1] y_{12}^{RO} and y_{12}^{LO} can also be extracted; however, since the system is passive, they should be equal to y_{21}^{RO} and y_{21}^{LO}, respectively. These relations can be used as a measurement verification.

$$Y_{right,int} = \begin{bmatrix} 0 & 0 \\ 0 & g_r \end{bmatrix}. \tag{168}$$

Then equations (163) and (164) become

$$\begin{bmatrix} Y_{11}^{LX} & Y_{12}^{LX} \\ Y_{21}^{LX} & Y_{22}^{LX} \end{bmatrix} = \frac{1}{(c_1c_2-1)^2} \begin{bmatrix} \frac{c_2}{d} & \frac{-1}{e} \\ \frac{-1}{d} & \frac{c_1}{e} \end{bmatrix} \begin{bmatrix} g_l & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \frac{c_2}{d} & \frac{-1}{d} \\ \frac{-1}{e} & \frac{c_1}{e} \end{bmatrix}, \tag{169}$$

$$\begin{bmatrix} Y_{11}^{RX} & Y_{12}^{RX} \\ Y_{21}^{RX} & Y_{22}^{RX} \end{bmatrix} = \frac{1}{(c_1c_2-1)^2} \begin{bmatrix} \frac{c_2}{d} & \frac{-1}{e} \\ \frac{-1}{d} & \frac{c_1}{e} \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 0 & g_r \end{bmatrix} \begin{bmatrix} \frac{c_2}{d} & \frac{-1}{d} \\ \frac{-1}{e} & \frac{c_1}{e} \end{bmatrix}. \tag{170}$$

Therefore, d and e are solved as

$$d = \frac{c_2}{c_1c_2-1}\sqrt{\frac{g_l}{Y_{11}^{LX}}}, \tag{171}$$

$$e = \frac{c_1}{c_1c_2-1}\sqrt{\frac{g_r}{Y_{22}^{LX}}}. \tag{172}$$

and Y_{ei}, Y_{ie}, and Y_{ii} are consequently solved. Hence, for a passive-system de-embedding, only OPEN, SHORT, LEFT, and RIGHT test structures are required in this four-port technique. This approach is thus feasible for accurate characterization of the substrate coupling in the millimeter-wave range.

CHAPTER VII

CONCLUSIONS AND FUTURE WORK

The contributions made by this work can be summarized as:

1. Implement 2-D device simulation in SiGe HBTs, then study and model the unique behaviors of SiGe HBTs such as the high-injection heterojunction barrier effect.

2. Develop calculation algorithms for systematic analysis of broadband noise and linearity in integrated circuits. Then develop an EDA tool that supports device- and circuit-level analysis, and supports configurable circuit optimization.

3. Present a new design methodology of SiGe LNA optimization and design rules-of-thumb.

4. Present a new two-port lumped-nonlinear-source behavioral model to simplify linearity analysis.

5. Present a new, simple four-port technique to accurately extract the *ac* and noise characteristics in the millimeter-wave range.

In the future, this work should be extended to explore the relation between the device technology and circuit performance in SiGe HBTs, especially regarding the impact of the device profile (e.g., doping density, Ge shape, etc.) and application environment (e.g., cryogenic temperature, thermal coupling, etc.) on the noise and linearity behaviors observed.

APPENDIX A

ANALYTICAL EXPRESSIONS OF LINEARITY IN LNA

Consider the simple model (Figure 86). The small-signal v_{in} and i_c satisfy the equation

$$v_s = B(s) \cdot \frac{v_{in}}{V_t} + A(s) \cdot i_c \tag{173}$$

where

$$A(s) = \left(s \cdot \tau_f + \frac{1}{\beta} \right) \cdot (Z_b + Z_e) + Z_e \tag{174}$$

$$B(s) = V_t \cdot [1 + s \cdot C_{te}(Z_b + Z_e)] \tag{175}$$

Z_b and Z_e are impedances shown in Figure 86.

$$Z_b = 50 + s \cdot l_b \tag{176}$$

$$Z_e = s \cdot l_e \tag{177}$$

where l_e and l_b are determined by the equations (85) and (86).

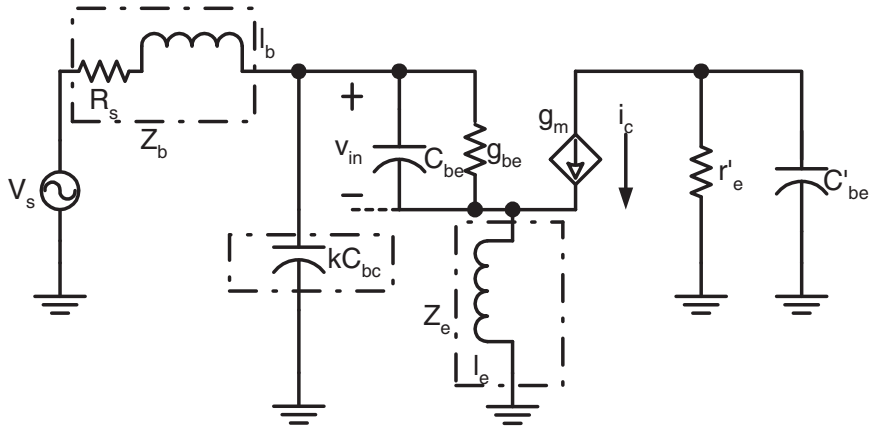

Figure 86: Simplified circuit for extracting l_e and l_b of cascode LNA, where k is a factor accounting for the Miller effect.

The Taylor expansion of i_C can be written as

$$i_c = \frac{I_C}{V_t} \cdot v_{in} + \frac{I_C}{2V_t^2} \cdot v_{in}^2 + \frac{I_C}{6V_t^3} \cdot v_{in}^3 \tag{178}$$

Substitute equation (178) into equation (173), and the 1st, 2nd, 3rd order of v_{in} and output voltage can be solved step by step [81] as

$$v_{in,1}(s) = \frac{v_s \cdot V_t}{B(s) + A(S) \cdot I_C} \tag{179}$$

$$\begin{aligned} v_{in,2}(s_1 + s_2) &= -C_{s1,s2} \cdot \frac{1}{V_t} \\ &\quad \times \frac{A(s_1 + s_2) \cdot I_C}{B(s_1 + s_2) + A(s_1 + s_2) \cdot I_C} \\ &\quad \times v_{in,1}(s1) \cdot v_{in,1}(s2) \end{aligned} \tag{180}$$

where

$$C_{s1,s2} = \begin{cases} \frac{1}{2} & , s1 = s2 \\ 1 & , s1 \neq s2 \end{cases} \tag{181}$$

$$v_{in,3}(2s_1 - s_2) = C(s_1, s_2) \cdot G(2s_1 - s_2) \cdot (1 - G(2s_1) - 2G(s_1 - s_2)) \cdot v_{in,1}(s_1) \tag{182}$$

where

$$C(s_1, s_2) = \frac{1}{6V_t^2} \cdot v_{in,1}(s_1) \cdot v_{in,1}(-s_2) \tag{183}$$

$$G(s) = \frac{A(s) \cdot I_C}{B(s) + A(s) \cdot I_C} \tag{184}$$

The output voltage can be then written as

$$v_{out}(2s_1 - s_2) = K(2s_1 - s_2) \cdot v_{in,3}(2s_1 - s_2) \cdot r_o \tag{185}$$

where

$$K(s) = \frac{B(s)}{A(s) \cdot V_t} \tag{186}$$

Since $2s_1 - s_2 \simeq s_1$, IM3 can be written as

$$IM3 \simeq \left| \frac{3}{4} \cdot C(s_1, s_2) \cdot L(s_1) \cdot (1 - G(2s_1) - 2G(s_1 - s_2)) \right| \tag{187}$$

where

$$L(s) = \frac{V_t \cdot K(s) \cdot G(s)}{I_C} \tag{188}$$

The cancellation comes from the third product term in the right side of equation (187).

APPENDIX B

TWO-PORT LNS BEHAVIOR MODEL ANALYSIS

B.1 Model transformation

In this section we demonstrate how to transform the nonlinear current source model (Y_1, Y_2, and Y_3) into other models (Z_1, Z_2, and Z_3; A_1, A_2, and A_3, etc.). Without loss of generality, any model defined above can be viewed as the behavior of the harmonic responses (including the fundamental frequency) to a fundamental stimulus. We denote the stimulus vector as S_{st} and the response vector as S_{res}, and the relation between these vectors and I/V vectors can be written as

$$S_{st} = AI + BV, \tag{189}$$

$$S_{res} = CI + DV, \tag{190}$$

where A, B, C, and D are $N \times N$ matrices for an N-port system. In most cases, these matrices are time-invariant. For example, in a current source model, the stimulus vector is $V = [v_1, v_2, ...v_N]^T$, and the response vector is $I = [i_1, i_2, ...i_N]^T$, and therefore $B = C$ equals the unit matrix E, and $A = D = 0$. In a voltage source model, the stimulus vector is I, and the response vector is V, and therefore $A = D = E$, $B = C = 0$.

As in Chapter III, we define two operators

$$\langle V(\omega_1), V(\omega_2) \mid Y_2(\omega_1, \omega_2)\rangle_\omega \text{ and } \langle V(\omega_1), V(\omega_2), V(\omega_3) \mid Y_3(\omega_1, \omega_2, \omega_3)\rangle_\omega$$

for brevity. The first operator transfers an $N \times N \times N$ tensor and two N-vectors into one N-vector, and is written as

$$\langle V(\omega_1), V(\omega_2) \mid Y_2(\omega_1, \omega_2)\rangle_\omega = \begin{bmatrix} i_{1,2nd}(\omega) \\ i_{2,2nd}(\omega) \\ \vdots \\ i_{N,2nd}(\omega) \end{bmatrix}, \tag{191}$$

where

$$i_{i,2nd}(\omega) = \sum_{(\omega_1,\omega_2)\in A_\omega} \left(\sum_{m=1}^{N} \left(\sum_{n=1}^{N} \left(y_{i,2}^{mn} \times v_m(\omega_1) v_n(\omega_2) \right) \right) \right), i = 1, 2...N. \tag{192}$$

The other operator transfers an $N \times N \times N \times N$ tensor and three N-vectors into one N-vector, and is written as

$$\langle V(\omega_1), V(\omega_2), V(\omega_3) \mid Y_3(\omega_1, \omega_2, \omega_3) \rangle_\omega = \begin{bmatrix} i_{1,3rd}(\omega) \\ i_{2,3rd}(\omega) \\ \vdots \\ i_{N,3rd}(\omega) \end{bmatrix}, \tag{193}$$

where

$$\begin{aligned} i_{i,3rd}(\omega) = \quad & \sum_{(\omega_1,\omega_2,\omega_3)\in B_\omega} \left\{ \sum_{m=1}^{N} \left(\sum_{n=1}^{N} \left(\sum_{j=1}^{N} \left(y_{i,3}^{mnj} \right. \right. \right. \right. \\ & \left. \left. \left. \left. \times v_m(\omega_1) v_n(\omega_2) v_j(\omega_3) \right) \right) \right) \right\}, \ i = 1, 2...N. \end{aligned} \tag{194}$$

To extract the behavioral model parameters (denoted by S_1, S_2, and S_3), one needs to retain $S_{st} = 0$ at harmonic frequencies (e.g., $\omega_1 + \omega_2$, $\omega_1 + \omega_2 + \omega_3$) and find the values of $S_{res}(\omega_1)$, $S_{res}(\omega_1 + \omega_2)$, and $S_{res}(\omega_1 + \omega_2 + \omega_3)$. For $S_{res}(\omega_1)$, the intermodulation is negligible compared with the linear portion. It can be written as

$$\begin{aligned} S_{res}(\omega_1) &= (CY_1(\omega_1) + D)V \\ &= (CY_1(\omega_1) + D)(AY_1(\omega_1) + B)^{-1} S_{st}(\omega_1) \\ &= S_1(\omega_1) S_{st}(\omega_1). \end{aligned} \tag{195}$$

Note that in some cases, some elements in Y_1 are equal to ∞. We set these elements to very large values to avoid a singularity.

For $S_{res}(\omega_1 + \omega_2)$, only the second intermodulation needs to be considered. Note that $S_{st}(\omega_1 + \omega_2) = 0$, and we have

$$[AY_1(\omega_1 + \omega_2) + B] V(\omega_1 + \omega_2) + A\langle V(\omega_1), V(\omega_2) \mid Y_2(\omega_1, \omega_2) \rangle_\omega = 0, \tag{196}$$

$$\Rightarrow V(\omega_1 + \omega_2) = -[AY_1(\omega_1 + \omega_2) + B]^{-1} \times A\langle V(\omega_1), V(\omega_2) \mid Y_2(\omega_1, \omega_2) \rangle_\omega. \tag{197}$$

Then,

$$
\begin{aligned}
S_{res}(\omega_1+\omega_2) &= [CY_1(\omega_1+\omega_2)+D]\,V(\omega_1+\omega_2) \\
&\quad + C\langle V(\omega_1), V(\omega_2) \mid Y_2(\omega_1,\omega_2)\rangle_\omega \\
&= \left\{C - [CY_1(\omega_1+\omega_2)+D]\,[AY_1(\omega_1+\omega_2)+B]^{-1}\,A\right\} \\
&\quad \times \langle V(\omega_1), V(\omega_2) \mid Y_2(\omega_1,\omega_2)\rangle_\omega \\
&= H3(\omega_1+\omega_2) \\
&\quad \times \langle H1(\omega_1)S_{st}(\omega_1), H1(\omega_2)S_{st}(\omega_2) \mid Y_2(\omega_1,\omega_2)\rangle_\omega \\
&= \langle S_{st}(\omega_1), S_{st}(\omega_2) \mid S_2(\omega_1,\omega_2)\rangle_\omega
\end{aligned} \tag{198}
$$

where

$$H1(\omega) = [AY_1(\omega)+B]^{-1}, \tag{199}$$

$$H2(\omega) = H1(\omega)A, \tag{200}$$

$$H3(\omega) = \{C - [CY_1(\omega)+D]\,H2(\omega)\}. \tag{201}$$

From the definition of the operators, one obtains

$$
\begin{aligned}
S_2(i,m,n) = \sum_{i'=1}^{N} \Bigg(H3_{ii'}(\omega_1+\omega_2) \sum_{m'=1}^{N} \Bigg(H1_{m'm}(\omega_1) \\
\times \sum_{n'=1}^{N} \Big(H1_{n'n}(\omega_2) Y_2(i',m',n') \Big) \Bigg) \Bigg).
\end{aligned} \tag{202}
$$

where X_{ij} denotes the element of matrix X in the i-th row and j-th column.

For $S_{res}(\omega_1+\omega_2+\omega_3)$, the second and third intermodulation needs to be considered. S_3 is derived in a similar manner

$$
\begin{aligned}
S_3(i,m,n,j) = \sum_{i'=1}^{N} \Bigg\{ H3_{ii'}(\omega_1+\omega_2+\omega_3) \sum_{m'=1}^{N} \Bigg(H1_{m'm}(\omega_1) \\
\times \sum_{n'=1}^{N} \Bigg(H1_{n'n}(\omega_2) \sum_{j'=1}^{N} \Big(H1_{j'j}(\omega_3) Y_3'(i',m',n',j') \Big) \Bigg) \Bigg) \Bigg\}.
\end{aligned} \tag{203}
$$

where

$$
\begin{aligned}
Y_3'(\omega_1,\omega_2,\omega_3) &= Y_3(\omega_1,\omega_2,\omega_3) \\
&\quad - Y_2'(\omega_1,\omega_2) \otimes Y_2(\omega_1+\omega_2,\omega_3)
\end{aligned}
$$

$$- Y_2'(\omega_2, \omega_3) \otimes Y_2(\omega_2 + \omega_3, \omega_1)$$
$$- Y_2'(\omega_1, \omega_3) \otimes Y_2(\omega_1 + \omega_3, \omega_2). \quad (204)$$

$$Y_2'(i, m, n) = \sum_{i'=1}^{N} H2_{ii'}(\omega_1 + \omega_2) Y_2(i', m, n). \quad (205)$$

For brevity, we introduce another operator $\otimes$, which combines two $N \times N \times N$ tensors into one $N \times N \times N \times N$ tensor. $Z = X \otimes Y$ is defined as

$$Z(i, m, n, j) = \sum_{k}^{N} X(k, m, n) Y(i, k, j). \quad (206)$$

This operator will be also used in the cascade model combination derivations.

B.2 Model combination

In this section we demonstrate how to combine two two-port behavioral models to a larger model. Parallel, series, and cascade combinations are considered. For a parallel combination, one can easily show that

$$Y_1'' = Y_1 + Y_1', \quad (207)$$
$$Y_2'' = Y_2 + Y_2', \quad (208)$$
$$Y_3'' = Y_3 + Y_3', \quad (209)$$

where Y_1, Y_2, and Y_3 are the current-source model parameters of one small system. Y_1', Y_2', and Y_3' are the current-source model parameters of the other small system, and Y_1'', Y_2'', and Y_3'' are the current-source model parameters of the combined system.

Similarly, for a series combination, one can obtain

$$Z_1'' = Z_1 + Z_1', \quad (210)$$
$$Z_2'' = Z_2 + Z_2', \quad (211)$$
$$Z_3'' = Z_3 + Z_3'. \quad (212)$$

where Z_1, Z_2, and Z_3 are the voltage-source model parameters of one small system, Z_1', Z_2', and Z_3' are the voltage-source model parameters of the other small system, and Z_1'', Z_2'', and Z_3'' are the voltage-source model parameters of the combined system.

For a cascade combination, the A_1, A_2, and A_3 model parameters are used. In this model form, the stimulus vector is $[i_1, V_1]^T$, and the response vector is $[-i_2, v_2]^T$. Then,

$$A_1'' = A_1' A_1. \tag{213}$$

Here, A_1, A_2, and A_3 are the model parameters of the first-stage. A_1', A_2', and A_3' are the model parameters of the second-stage, and A_1'', A_2'', and A_3'' are the model parameters of the overall system.

A_2'' is written as

$$\begin{aligned} A_2''(i,m,n) = &\sum_{m'=1}^{N} \left(A_{1,m'm}(\omega_1) \sum_{n'=1}^{N} \left(A_{1,n'n}(\omega_2) A_2'(i,m',n') \right) \right) \\ &+ \sum_{i'=1}^{N} \left(A_{1,ii'}'(\omega_1+\omega_2) A_2(i',m,n) \right). \end{aligned} \tag{214}$$

A_3'' is written as

$$\begin{aligned} A_3''(i,m,n,j) = &\sum_{m'=1}^{N} \Bigg(A_{1,m'm}(\omega_1) \sum_{n'=1}^{N} \Bigg(A_{1,n'n}(\omega_2) \\ &\times \sum_{j'=1}^{N} \left(A_{1,j'j}(\omega_3) A_3'(i,m',n',j') \right) \Bigg) \Bigg) \\ &+ \sum_{i'=1}^{N} \left(A_{1,ii'}'(\omega_1+\omega_2+\omega_3) A_3(i,m',n',j') \right) \\ &+ A_3^C(i,m,n,j). \end{aligned} \tag{215}$$

where

$$\begin{aligned} A_3^C(\omega_1,\omega_2,\omega_3) = &A_2(\omega_1,\omega_2) \otimes A_2'(\omega_1+\omega_2,\omega_3) \\ &+ A_2(\omega_2,\omega_3) \otimes A_2'(\omega_2+\omega_3,\omega_1) \\ &+ A_2(\omega_1,\omega_3) \otimes A_2'(\omega_1+\omega_3,\omega_2) \end{aligned} \tag{216}$$

is the cross-intermodulation term.

B.3 Impact of harmonic impedance

In this section we derive the *IM2* and *IM3* as functions of the source/load harmonic admittance.

At the fundamental frequency, the voltages at the input and output are written as

$$v_{in}(\omega_1) = k_{in}(\omega_1) I_a, \tag{217}$$

$$v_{out}(\omega_1) = k_{out}(\omega_1) I_a, \tag{218}$$

where I_a is the value of the current source, and

$$k_{in}(\omega) = \frac{y_{22}(\omega) + y_L(\omega)}{\Delta(\omega) + B(\omega)}, \tag{219}$$

$$k_{out}(\omega) = \frac{-y_{21}(\omega)}{\Delta(\omega) + B(\omega)}. \tag{220}$$

y_{11}, y_{12}, y_{21}, and y_{22} are the y-parameters (elements of Y_1), and y_S and y_L are the source and load admittances, respectively.

$$\Delta(\omega) = y_{11}(\omega)y_{22}(\omega) - y_{12}(\omega)y_{21}(\omega), \tag{221}$$

$$B(\omega) = y_{22}(\omega)y_S(\omega) + y_{11}(\omega)y_L(\omega) + y_S(\omega)y_L(\omega). \tag{222}$$

Then, the second-order current source $i_{1,2nd}$ and $i_{2,2nd}$ are

$$i_{1,2nd}(2\omega_1) = m_{in}(\omega_1, \omega_1)I_a^2, \tag{223}$$

$$i_{2,2nd}(2\omega_1) = m_{out}(\omega_1, \omega_1)I_a^2. \tag{224}$$

$$\begin{aligned} m_{in}(\omega_1, \omega_1) &= y_{1,2}^{11}(\omega_1, \omega_1)k_{in}(\omega_1)^2 + y_{1,2}^{22}(\omega_1, \omega_1)k_{out}(\omega_1)^2 \\ &+ [y_{1,2}^{12}(\omega_1, \omega_1) + y_{1,2}^{21}(\omega_1, \omega_1)]k_{in}(\omega_1)k_{out}(\omega_1), \end{aligned} \tag{225}$$

$$\begin{aligned} m_{out}(\omega_1, \omega_1) &= y_{2,2}^{11}(\omega_1, \omega_1)k_{in}(\omega_1)^2 + y_{2,2}^{22}(\omega_1, \omega_1)k_{out}(\omega_1)^2 \\ &+ [y_{2,2}^{12}(\omega_1, \omega_1) + y_{2,2}^{21}(\omega_1, \omega_1)]k_{in}(\omega_1)k_{out}(\omega_1). \end{aligned} \tag{226}$$

The output and input voltage at the second-order harmonic $2\omega_1$ are

$$v_{out}(2\omega_1) = \frac{A(2\omega_1) + i_{2,2nd}(2\omega_1)y_S(2\omega_1)}{\Delta(2\omega_1) + B(2\omega_1)}, \tag{227}$$

$$v_{in}(2\omega_1) = \frac{D(2\omega_1) + i_{1,2nd}(2\omega_1)y_L(2\omega_1)}{\Delta(2\omega_1) + B(2\omega_1)}, \tag{228}$$

where

$$A(2\omega_1) = y_{11}(2\omega_1)i_{2,2nd}(2\omega_1) - y_{21}(2\omega_1)i_{1,2nd}(2\omega_1), \tag{229}$$

$$D(2\omega_1) = y_{22}(2\omega_1)i_{1,2nd}(2\omega_1) - y_{12}(2\omega_1)i_{2,2nd}(2\omega_1). \tag{230}$$

Then,

$$\begin{aligned} IM2 &= \frac{|v_{out}(2\omega_1)|^2}{|v_{out}(\omega_1)|^2}\frac{real(y_L(2\omega_1))}{real(y_L(\omega_1))} \\ &= A_{L2}\frac{real(y_L(2\omega_1))}{|y_L(2\omega_1) + C_L|^2} \\ &= A_{S2}\frac{|y_S(2\omega_1) + B_{S2}|^2}{|y_S(2\omega_1) + C_S|^2}. \end{aligned} \tag{231}$$

where

$$A_{L2} = \frac{|A(2\omega_1) + i_{2,2nd}(2\omega_1)y_S(2\omega_1)|^2}{|k_{out}(\omega_1)[y_{11}(2\omega_1) + y_S(2\omega_1)]|^2 I_a^2 Real(y_L(\omega_1))}, \tag{232}$$

$$C_L = \frac{D(2\omega_1) + y_{22}(2\omega_1)y_S(2\omega_1)}{y_{11}(2\omega_1) + y_S(2\omega_1)}, \tag{233}$$

and

$$A_{S2} = \frac{|i_{2,2nd}(2\omega_1)|^2 Real(y_L(2\omega_1))}{|k_{out}(\omega_1)[y_{22}(2\omega_1) + y_L(2\omega_1)]|^2 I_a^2 Real(y_L(\omega_1))}, \tag{234}$$

$$B_{S2} = A(2\omega_1)/i_{2,2nd}(2\omega_1), \tag{235}$$

$$C_S = \frac{D(2\omega_1) + y_{11}(2\omega_1)y_L(2\omega_1)}{y_{22}(2\omega_1) + y_L(2\omega_1)}. \tag{236}$$

The nonlinear current source at the $2\omega_1 - \omega_2$ includes three terms: the current generated from the third-order intermodulation (ω_1, ω_1, $-\omega_2$ intermodulation), the current generated from the intermodulation between ω_1 and $\omega_1 - \omega_2$, and the current generated from the intermodulation between $2\omega_1$ and $-\omega_2$. Only the last term is a function of the source/load second-harmonic admittance. Let us denote $I^C_{1,3rd}$ and $I^C_{2,3rd}$ as the sum of the first two terms, and $I^V_{1,3rd}$ and $I^V_{2,3rd}$ as the last term at the input and output, respectively. The output voltage at $2\omega_1 - \omega_2$ is

$$\begin{aligned} v_{out}(2\omega_1 - \omega_2) = {} & k_{out}(2\omega_1 - \omega_2)\left(I^C_{1,3rd} + I^V_{1,3rd}\right) \\ & + k_{ref}(2\omega_1 - \omega_2)\left(I^C_{2,3rd} + I^V_{2,3rd}\right), \end{aligned} \tag{237}$$

where

$$k_{ref}(\omega) = y_{22}(\omega) + y_L(\omega) - \frac{y_{12}(\omega)y_{21}(\omega)}{y_{11}(\omega) + y_S(\omega)}. \tag{238}$$

Since $I^V_{1,3rd}$ and $I^V_{2,3rd}$ is generated from the intermodulation between $2\omega_1$ and $-\omega_2$, for brevity, they are written as

$$\begin{bmatrix} I^V_{1,3rd} \\ \\ I^V_{2,3rd} \end{bmatrix} = \langle V(2\omega_1), V(-\omega_2) \mid Y_2(2\omega_1, -\omega_2)\rangle_{2\omega_1 - \omega_2}, \tag{239}$$

where $V(\omega) = [v_{in}(\omega), v_{out}(\omega)]^T$. Taking equations (227), (228) and (239) into equation (237), one derives $v_{out}(2\omega_1 - \omega_2)$ in the form of

$$v_{out}(2\omega_1 - \omega_2) = \frac{G + Ey_S(2\omega_1) + Fy_L(2\omega_1)}{\Delta(2\omega_1) + B(2\omega_1)}, \tag{240}$$

where G, E, and F are intermediate terms derived using the equations above, and are not functions of $y_S(2\omega_1)$ or $y_L(2\omega_1)$. *IM3* can be written as

$$\begin{aligned} IM3 &= \frac{|v_{out}(2\omega_1 - \omega_2)|^2}{|v_{out}(\omega_1)|^2} \frac{Real(y_L(2\omega_1 - \omega_2))}{Real(y_L(\omega_1))} \\ &= A_{L3} \frac{|y_L(2\omega_1) + B_{L3}|^2}{|y_L(2\omega_1) + C_L|^2} \\ &= A_{S3} \frac{|y_S(2\omega_1) + B_{S3}|^2}{|y_S(2\omega_1) + C_S|^2}. \end{aligned} \tag{241}$$

where

$$A_{L3} = \frac{|F|^2 Real(y_L(2\omega_1 - \omega_2))}{|k_{out}(\omega_1)[y_{11}(2\omega_1) + y_S(2\omega_1)]|^2 I_a^2 Real(y_L(\omega_1))}, \tag{242}$$

$$B_{L3} = \frac{G + E y_S(2\omega_1 - \omega_2)}{F}, \tag{243}$$

and

$$A_{S3} = \frac{|E|^2 Real(y_L(2\omega_1 - \omega_2))}{|k_{out}(\omega_1)[y_{22}(2\omega_1) + y_L(2\omega_1)]|^2 I_a^2 Real(y_L(\omega_1))}, \tag{244}$$

$$B_{S3} = \frac{G + F y_L(2\omega_1 - \omega_2)}{E}. \tag{245}$$

Expressions for *IM2* at $\omega_2 - \omega_1$ and *IM3* as functions of source/load IF impedance are derived in the same manner.

APPENDIX C

FOUR-PORT DE-EMBEDDING TECHNIQUE

C.1 The four-port noise correlation matrix de-embedding

Substituting $I_i = -Y^{INT}V_i$ into the equation (160), one obtains

$$I_e + \widetilde{i_{n,e}} - Y_{ei}(Y^{INT} + Y_{ii})^{-1}(\widetilde{i_{n,i}} + \widetilde{i_{n,int}}) = [Y_{ee} - Y_{ei}(Y^{INT} + Y_{ii})^{-1}Y_{ie}]V_e, \quad (246)$$

$$\Rightarrow \widetilde{i_{n,total}} = \widetilde{i_{n,e}} - Y_{ei}(Y^{INT} + Y_{ii})^{-1}(\widetilde{i_{n,i}} + \widetilde{i_{n,int}}). \quad (247)$$

where $i_{n,total}$ are the equivalent noise current sources at the extrinsic ports. The resultant noise correlation is

$$\begin{aligned} SY_{n,total} &= \widetilde{i_{n,total}}\widetilde{i^*_{n,total}} \\ &= [\widetilde{i_{n,e}} - Y_{ei}(Y^{INT} + Y_{ii})^{-1}(\widetilde{i_{n,i}} + \widetilde{i_{n,int}})] \times [\widetilde{i_{n,e}} - Y_{ei}(Y^{INT} + Y_{ii})^{-1}(\widetilde{i_{n,i}} + \widetilde{i_{n,int}})]^*. \end{aligned} \quad (248)$$

Note that the intrinsic device noise sources and the parasitic noise sources are un-correlated, and thus one obtains

$$\begin{aligned} SY_{n,total} &= \widetilde{i_{n,e}}\widetilde{i^*_{n,e}} + Y_{ei}(Y^{INT} + Y_{ii})^{-1}(\widetilde{i_{n,i}}\widetilde{i^*_{n,i}} + \widetilde{i_{n,int}}\widetilde{i^*_{n,int}})[(Y^{INT} + Y_{ii})^{-1}]^*Y^*_{ei} \\ &\quad - Y_{ei}(Y^{INT} + Y_{ii})^{-1}\widetilde{i_{n,i}}\widetilde{i^*_{n,e}} - \widetilde{i^*_{n,i}}\widetilde{i_{n,e}}[(Y^{INT} + Y_{ii})^{-1}]^*Y^*_{ei} \\ &= SY_{n,ee} + Y_T SY_{n,ii}Y^*_T + Y_T SY_{n,int}Y^*_T - Y_T SY_{n,ie} - SY_{n,ei}Y^*_T. \end{aligned} \quad (249)$$

where

$$Y_T = Y_{ei}(Y^{INT} + Y_{ii})^{-1}, \quad (250)$$

$$Y_{n,ee} = 4kTReal(Y_{ee}), \quad (251)$$

$$Y_{n,ei} = 4kTReal(Y_{ei}), \quad (252)$$

$$Y_{n,ie} = 4kTReal(Y_{ie}), \quad (253)$$

$$Y_{n,ii} = 4kTReal(Y_{ii}). \quad (254)$$

Hence, the intrinsic noise current correlation matrix can be calculated as

$$SY_{n,int} = (Y_T)^{-1}(SY_{n,total} - SY_{n,ee})(Y_T^*)^{-1} - SY_{n,ii} + (Y_T)^{-1}SY_{n,ei} + SY_{n,ie}(Y_T^*)^{-1}. \quad (255)$$

C.2 Linear expansion of the y-parameters of the test structures

The $Y_{test,int} + Y_{ii}$ for the OPEN, LEFT, RIGHT, THROUGH test structures are

$$Y_{open,int} + Y_{ii} = \begin{bmatrix} Y_{ii,11} & Y_{ii,12} \\ Y_{ii,21} & Y_{ii,22} \end{bmatrix}, \quad (256)$$

$$Y_{left,int} + Y_{ii} = \begin{bmatrix} Y_{ii,11} + g_l & Y_{ii,12} \\ Y_{ii,21} & Y_{ii,22} \end{bmatrix}, \quad (257)$$

$$Y_{right,int} + Y_{ii} = \begin{bmatrix} Y_{ii,11} & Y_{ii,12} \\ Y_{ii,21} & Y_{ii,22} + g_r \end{bmatrix}, \quad (258)$$

$$Y_{through,int} + Y_{ii} = \begin{bmatrix} Y_{ii,11} + A & Y_{ii,12} - A \\ Y_{ii,21} - A & Y_{ii,22} + A \end{bmatrix}. \quad (259)$$

Consider $(Y_{open,int} + Y_{ii})^{-1}$ and $(Y_{left,int} + Y_{ii})^{-1}$

$$(Y_{open,int} + Y_{ii})^{-1} = \begin{bmatrix} Y_{ii,22}/D & -Y_{ii,12}/D \\ -Y_{ii,21}/D & Y_{ii,11}/D \end{bmatrix}, \quad (260)$$

$$(Y_{left,int} + Y_{ii})^{-1} = \begin{bmatrix} Y_{ii,22}/(D + g_l Y_{ii,22}) & -Y_{ii,12}/(D + g_l Y_{ii,22}) \\ -Y_{ii,21}/(D + g_l Y_{ii,22}) & (Y_{ii,11} + g_l)/(D + g_l Y_{ii,22}) \end{bmatrix}, \quad (261)$$

where $D = |Y_{ii}| = Y_{ii,11}Y_{ii,22} - Y_{ii,12}Y_{ii,21}$. If we choose $x = D/(D + g_l Y_{ii,22})$, then

$$\begin{aligned} Y^{LEFT} - xY^{OPEN} - (1-x)Y^{SHORT} &= Y_{ei}[(Y_{left,int} + Y_{ii})^{-1} - x(Y_{open,int} + Y_{ii})^{-1}]Y_{ie} \\ &= Y_{ei}\begin{bmatrix} 0 & 0 \\ 0 & a \end{bmatrix} Y_{ie}. \end{aligned} \quad (262)$$

where $a = g_l/(D + g_l Y_{ii,22})$. Thus $x = D/(D + g_l Y_{ii,22})$ is one solution of the equation

$$Y^{LEFT} - xY^{OPEN} - (1-x)Y^{SHORT} = 0. \quad (263)$$

On the other hand, since

$$\begin{vmatrix} Y_{ii,22}[1/(D + g_l Y_{ii,22}) - 1/D] & -Y_{ii,12}/[1/(D + g_l Y_{ii,22}) - 1/D] \\ -Y_{ii,21}/[1/(D + g_l Y_{ii,22}) - 1/D] & Y_{ii,11}/[1/(D + g_l Y_{ii,22}) - 1/D] + g_l/(D + g_l Y_{ii,22}) \end{vmatrix} = 0,$$

$$\Rightarrow \left| Y^{LEFT} - Y^{OPEN} \right| = 0. \tag{264}$$

Hence, $D/(D + g_l Y_{ii,22})$ and 1 are the only two solutions of equation (151). Similarly one can prove that $D/(D + g_l Y_{ii,11})$ and 1 are the only two solutions of equation (152), and that $D/(D + A(Y_{ii,11} + Y_{ii,12} + Y_{ii,21} + Y_{ii,22}))$ and 1 are the only two solutions of equation (153). The values of x, y and z are thus fully determined.

REFERENCES

[1] H. Kroemer, "Theory of a wide-gap emitter for transistors," *Proceedings of IRE*, vol. 45, pp. 1535-1537, 1957.

[2] B.S. Meyerson, "Low-temperature silicon epitaxy by ultrahigh vacuum/chemical vapor deposition," *Appl. Phys. Lett.*, vol. 48, pp. 797-799, 1986.

[3] S.S. Iyer, G.L. Patton, S.L. Delage, S. Tiwari, and J.M.C. Stork , "Silicon-germanium base heterojunction bipolar transistors by molecular beam epitaxy," *Technical Digest of IEEE International Electron Devices Meeting*, pp. 874-876, 1987.

[4] J.D. Cressler, "SiGe HBT technology: A new contender for Si-based RF and microwave circuit applications," *IEEE Trans. Microwave Theory and Tech.*, vol. 46, pp. 572-589, 1998.

[5] A.J. Joseph, D. Coolbaugh, D.L. Harame, G. Freeman, S. Subbanna, M. Doherty, J. Dunn, C. Dickey, D. Greenberg, R. Groves, M. Meghelli, A. Rylyakov, M. Sorna, O. Schreiber, D. Herman, and T. Tanji, "0.13 μm 210 GHz f_T SiGe HBTs - expanding the horizons of SiGe BiCMOS," *Technical Digest of IEEE International Solid-State Circuits Conference*, pp. 180-182, 2002.

[6] K. Oda, E. Ohue, I. Suzumura, R. Hayami, A. Kodama, H. Shimamoto, and K. Washio, "Self-aligned selective-epitaxial-growth $Si_{1-x-y}Ge_xC_y$ HBT technology featuring 170-GHz f_{max}," *Technical Digest of IEEE International Electron Devices Meeting*, pp. 332-335, 2001.

[7] M. Racanelli, K. Schuegraf, A. Kalburge, A. Kar-Roy, B. Shen, C. Hu, D. Chapek, D. Howard, D. Quon, F. Wang, G. U'ren, L. Lao, H. Tu, J. Zheng, J. Zhang, K. Bell, K. Yin, P. Joshi, S. Akhtar, S. Vo, T. Lee, W. Shi, and P. Kempf, "Ultra high speed SiGe npn for advanced BiCMOS technology," *Technical Digest of IEEE International Electron Device Meeting*, pp. 15.3.1-15.3.4, 2001.

[8] J.D. Cressler and G. Niu, *Silicon Germanium Heterojunction Bipolar Transistors*, Boston, MA: Artech House, 2003.

[9] A.J. Joseph, J.D. Cressler, D.M. Richey, and G. Niu, "Optimization of SiGe HBTs for operation at high current densities," *IEEE Trans. Electron Devices*, vol. 46 pp. 1347-1356, 1999.

[10] Q. Liang, J.D. Cressler, G. Niu, R.M. Malladi, K. Newton, and D.L. Harame, "A high-injection transit-time model for heterojunction barrier effects in SiGe HBTs," *Proceedings of IEEE Bipolar/BiCMOS Circuits and Technology Meeting*, pp. 236-238, 2002.

[11] M. Pfost, H-M. Rein, and T. Holzwarth, "Modeling substrate effects in the design of high-speed Si-Bipolar IC's," *IEEE J. Solid-State Circuits*, vol. 31, pp. 1493-1501, Oct. 1996.

[12] J.S. Rieh, D. Greenberg, B. Jagannathan, G. Freeman, and S. Subbanna, "Measurement and modeling of thermal resistance of high speed SiGe heterojunction bipolar transistors," *Proceedings of IEEE Bipolar/BiCMOS Circuits and Technology Meeting*, pp. 110-113 2001.

[13] S.F. Shams, C.C. McAndrew, L. Ik-Sung, and A. Zlotnicka, "SiGe HBT self-heating modeling and characterization from AC Data," *Proceedings of IEEE Bipolar/BiCMOS Circuits and Technology Meeting*, pp. 92-95, 2002.

[14] T. Pesic, N. Jankovic, and J. Karamarkovic, "Modeling of the inverse base width modulation effect in SiGe base HBT for circuit simulation," *The Fourth International Conference on Advanced Semiconductor Devices and Microsystems*, pp. 187-190, 2002.

[15] W. Shockley, "The impedance field method of noise calculation in active semiconductor devices" *in: Quantum Theory of Atoms, Molecules, and the Solid-State.*, P.-O. Lowdin, Ed. New York: Academic, 1966.

[16] A. Van der Ziel, *Noise in Solid State Devices and Circuits*, New York, NY: Wiley, 1968.

[17] J. Lee, L. Geunho, G. Niu, J.D. Cressler, J.H. Kim, J.C. Lee, B. Lee, and N.Y. Kim, "The design of SiGe HBT LNA for IMT-2000 mobile application," *IEEE MTT-S International Microwave Symposium Digest*, pp. 1261-1264, 2002.

[18] O. Shana'a, I. Linscott, and L. Tyler, "Frequency-scalable SiGe bipolar RFIC front-end design," *Proceedings of IEEE Custom Integrated Circuits Conference (CICC)*, pp. 21-24, 2000.

[19] S. Hackl, T.F. Meiter, M. Wurzer, H. Knapp, K. Aufinger, L. Treitinger, and A.L. Scholtz, "Low-noise, low-power monolithically integrated active 20 GHz mixer in SiGe technology," *Electronics Letters*, vol. 37, no. 1 pp. 36-37, Jan. 2001.

[20] W. Durr, U. Erben, A. Schuppen, H. Dietrich, and H. Schumacher, "Low-power low-noise active mixers for 5.7 and 11.2 GHz using commercially available SiGe HBT MMIC technology," *Electronics Letters*, vol. 34, pp. 1994-1996, Oct. 1998.

[21] X. Wang, D. Wang, K. Schelkle, and P. Bacon, "Fully integrated low phase noise VCO design in SiGe BiCMOS technology," *IEEE Radio and Wireless Conference*, pp. 109-112, 2001.

[22] B. Razavi, *RF Microelectronics*, Indianapolis, IN: Prentice Hall, 1998.

[23] M. Leach, *Course Notes*, ece6416 Georgia Institute of Technology, Fall 2002.

[24] *ATN-NP5 Noise System Manual 2001*.

[25] M.C.A.M. Koolen, J.A.M. Geelen, and M.P.J.G. Versleijen, "An improved de-embedding technique for on-wafer high-frequency characterization," *Proceedings of IEEE Bipolar/BiCMOS Circuits and Technology Meeting*, pp. 188-191, Sept. 1991.

[26] E.P. Vandamme, D.M.M.P. Schreurs, and C.V. Dinther, "Improved three-step de-embedding method to accurately account for the influence of pad parasitics in silicon on-wafer RF test-structures," *IEEE Trans. Electron Devices*, vol. 48, pp. 737-742, Apr. 2001.

[27] C.H. Chen and M.J. Deen, "A general noise and s-parameter de-embedding procedure for on-wafer high-frequency noise measurements of MOSFETs," *IEEE Tran. Microwave Theory and Tech.*, vol. 49, No. 5, pp. 1004-1005, May 2001.

[28] S. Bousnina, C. Falt, P. Mandeville, A.B. Kouki, and F.M. Ghannouchi, "An accurate on-wafer de-embedding technique with application to HBT device characterization," *IEEE Tran. Microwave Theory and Tech.*, vol. 50, No. 2, pp. 420-424, Feb. 2002.

[29] G. Niu, Z. Jin, J.D. Cressler, R. Rapeta, A.J. Joseph, and D. Harame, "Transistor noise in SiGe HBT RF technology," *IEEE J. Solid-State Circuits*, vol. 36, pp. 1424-1427, Sept. 2001.

[30] E.M. Bailey, "Steady state harmonic analysis of nonlinear networks," *PH.D. dissertation*, Standford Univ., Stanford, CA, 1968.

[31] J.C. Lindenlaub, "An approach for finding the sinusoidal steady state response of nonlinear systems," *in Proc. 7th Ann. Allerton Conf. Circuit and System Theory* (Univ. Illinois, Chicago), 1969.

[32] S. Narayanan, "Transistor distortion analysis using volterra series representation," *Bell System Technical Journal*, vol. 46, no.3, pp.991-1024, 1967.

[33] V. Aparin, E. Zeisel, and P. Gazzerro, "Highly linear SiGe BiCMOS LNA and mixer for cellular CDMA/AMPS applications," *IEEE Radio Frequency Integrated Circuits (RFIC) Symposium*, pp. 129-132, 2002.

[34] P. Juurakko, V. Saari, J. Ryynanen, and K. Halonen, "Differential SiGe power amplifier for 3GPP WCDMA," *IEEE Radio Frequency Integrated Circuits (RFIC) Symposium*, pp. 133-136, 2003.

[35] P. Wambacq and W. Sansen, *Distortion Analysis of Analog Integrated Circuits*, New York, NY: Kluwer Academic, 1998.

[36] G. Niu, Q. Liang, J.D. Cressler, C.S. Webster, and D.L. Harame. "RF linearity characteristics of SiGe HBTs," *IEEE Trans. Microwave Theory Tech.*, vol. 49, pp. 1558-1565, 2001.

[37] K.L. Fong, "High-frequency analysis of linearity improvement technique of common-emitter transconductance stage using a low-frequency-trap network," *IEEE J. Solid-State Circuits*, vol. 35, pp. 1249-1252, 2000.

[38] L. Sheng and L.E. Larson, "A Si/SiGe BiCMOS mixer with 3rd-order nonlinearity cancellation for WCDMA applications" *IEEE Radio Frequency Integrated Circuits (RFIC) Symposium*, pp. 1551-1554, 2003.

[39] Q. Liang, G. Niu, J.D. Cressler, S. Taylor, and D.L. Harame, "On the optimization and design of SiGe HBT cascode low-noise amplifiers," *Solid-State Electronics*, accepted 2004.

[40] Q. Liang, J.M. Andrews, J.D. Cressler, and G. Niu, "General analysis of the impact of harmonic impedance on linearity in SiGe HBTs," *Proceedings of IEEE Bipolar/BiCMOS Circuits and Technology Meeting*, pp. 48-51, 2004.

[41] A.J. Joseph, "The physics, optimization, and modeling of cryogenically operated silicon-germanium heterojunction bipolar transistors," *Ph.D. dissertation*, Auburn, AL: Auburn University, 1997.

[42] S.L. Salmon, J.D. Cressler, R.C. Jaeger, and D.L. Harame, "The influence of Ge grading on the bias and temperature characteristics of SiGe HBT's for precision analog circuits," *IEEE Trans. Electron Devices*, vol. 47, pp. 292-298, Feb., 2002.

[43] Q. Liang, J.D. Cressler, G. Niu, R.M. Malladi, K. Newton, and D.L. Harame, "A physics-based high-injection transit-time model applied to barrier effects in SiGe HBTs," *IEEE Trans. Electron Devices*, vol. 49, no. 10, pp. 1807-1813, Oct. 2002.

[44] C.T. Kirk, "Theory of transistor cutoff frequency falloff at high current densities," *IRE Trans. Electron Devices*, vol. 3, no. 3, pp. 164-170, Mar. 1962.

[45] SYNOPSYS, *MEDICI: Two-dimensional Device Simulation Program User Manual*, version 2002.4, Feb., 2003.

[46] C.C. McAndrew, J.A. Seitchik, D.F. Bowers, M. Dunn, M. Foisy, I. Getreu, M. McSwain, S. Moinian, J. Parker, D.J. Roulston, M. Schröter, P. Van Wijnen, and L.F. Wagner, "VBIC95, the vertical bipolar inter-company model," *IEEE J. Solid-State Circuits*, vol. 31, pp. 1476-1483, Oct. 1996.

[47] M. Schröter and L. Tzung-Yin "Physics-based minority charge and transit time modeling for bipolar transistors," *IEEE Trans. Electron Devices*, vol. 46, pp. 288-300, Feb. 1999.

[48] S. Tiwari and D.J. Frank, "Analysis of the operation of GaAlAs/GaAs HBT's," *IEEE Trans. Electron Devices*, vol. 36, pp. 2105-2121, Oct. 1989.

[49] Z.P. Yu, P.E. Cottrell, and R.W. Dutton, "Modeling and simulation of high-level injection behavior in double heterojunction transistors," *Proceedings of IEEE Bipolar/BiCMOS Circuits and Technology Meeting*, pp. 192-194, 1990.

[50] D.R. Greenberg, M. Rivier, P. Girard, E. Bergeault, J. Moniz, D. Ahlgren, G. Freeman, S. Subbanna, S.J. Jeng, K. Stein, D. Nguyen-Ngoc, K. Schonenberg, J. Malinowski, D. Colavito, D.L. Harame, and B. Meyerson, "Large-signal performance of high-BV_{CEO} graded epi-base SiGe HBTs at wireless frequencies," *Technical Digest of IEEE International Electron Devices Meeting*, pp. 799-802, 1997.

[51] A.J. Joseph, J.D. Cressler, and D.M. Richey, "Operation of SiGe heterojunction bipolar transistors in the liquid-helium temperature regime," *IEEE Electron Device Letters*, vol. 16, pp. 268-270, June 1995.

[52] H. Kroemer, "Two integral relations pertaining to the electron transport through a bipolar transistor with a nonuniform energy gap in the base region," *Solid-State Electronics*, vol. 28, pp. 1101-1103, 1985.

[53] D.E. Root, J. Wood, and N. Tufillaro "New techniques for non-linear behavioral modeling of microwave/RF ICs from simulation and nonlinear microwave measurements," *IEEE Design Automation Conference*, pp. 85-90, Jun. 2003.

[54] D.D. Weiner and G.H. Naditch, "A scattering variable approach to the volterra analysis of nonlinear systems," *IEEE Trans. Microwave Theory and Techniques,* vol. 24, pp. 422-433, Jul. 1976.

[55] A. Saleh, "Matrix analysis of mildly nonlinear multiple-input, multiple-output systems with memeory," *The Bell Labs Technical Journal,* vol. 61, pp. 2221-2243, 1982.

[56] F. Verbeyst and M.V. Bossche, "VIOMAP, the s-parameter equivalent for weakly nonlinear RF and microwave devices," *IEEE Trans. Microwave Theory and Techniques,* vol. 42, pp. 2531-2535, Dec. 1994.

[57] J.C. Pedro, J.C. Madaleno, and J.A. García, "Theoretical basis for extraction of mildly nonlinear behavioral models," *International Journal of RF and Microwave Computer-Aided Engineering,* vol. 13, pp. 40-53, Jan. 2003.

[58] J.C. Pedro and N.B. Carvalho, *Intermodulation Distortion in Microwave and Wireless Circuits,* Artech House, London, 2003.

[59] G. Gielen and W. Sansen, *Symbolic Analysis for Automated Design of Analog Integrated Circuits*, New York, NY: Kluwer, 1991.

[60] M. Sanden, O. Marinov, M.J. Deen, and M. Ostling "A new model for the low-frequency noise and the noise level variation in polysilicon emitter BJTs," *IEEE Trans. Electron Devices*, vol. 49, pp. 514-520, Mar. 2002.

[61] H. Hillbrand and P.H. Russer, "An efficient method for computer-aided noise analysis of linear amplifier networks," *IEEE Trans. Circuit and Systems*, vol. 23, pp. 235-238, Apr. 1976.

[62] F. Bonani, G. Ghione, M.R. Pinto, and R.K. Smith, "An efficient approach to noise analysis through multidimensional physics-based models," *IEEE Trans. Electron Devices*, vol. 45, pp. 261-269, 1998.

[63] Agilent, *Verilog-A Reference Manual*, http://eesof.tm.agilent.com/docs/rfdedoc2003C.

[64] P. Crippa, S. Orcioni, F. Ricciardi, and C. Turchetti, "Design of a 4.4 to 5 GHz LNA in 0.25-μm SiGe BiCMOS technology," *IEEE International Symposium on Circuits and Systems*, pp. 333-336, 2003.

[65] Q. Liang, G. Niu, J.D. Cressler, S. Taylor, and D.L. Harame, "Geometry and bias current optimization for SiGe HBT cascode low-noise amplifiers," *IEEE Radio Frequency Integrated Circuits (RFIC) Symposium*, pp. 407-410, 2002.

[66] M. Soyuer, J-O. Plouchart, H. Ainspan, and J. Burghartz, "A 5.8-GHz 1-V low-noise amplifier in SiGe bipolar technology," *IEEE Radio Frequency Integrated Circuits (RFIC) Symposium*, pp. 19-22, 1997.

[67] H.A. Haus, W.R. Atkinson, W.H. Fonger, W.W. Mcleod, G.M. Branch, W.A. Harris, E.K. Stodola, W.B. Davenport, Jr.S.W. Harrison, and T.E. Talpey, "Representation of noise in linear twoports," *Proceedings of IRE*, vol. 48, pp. 67-74, Jan. 1960.

[68] S. Zhang, G. Niu, J.D. Cressler, A.J. Joseph, G. Freeman, and D.L. Harame, "The effects of geometrical scaling on the frequency response and noise performance of SiGe HBTs," *IEEE Trans. Electron Devices.*, vol. 49, pp. 429-435, Mar. 2002.

[69] G. Niu, J.D. Cressler, S. Zhang, U. Gogineni, and D.C. Ahlgren, "Measurement of collector-base junction avalanche multiplication effects in advanced UHV/CVD SiGe HBT," *IEEE Trans. Electron Devices*, vol. 46, pp. 1007-1015, May 1999.

[70] H. Schumacher, U. Erben, W. Dürr, and K-B. Schad, "Low-noise, low-power wireless front-end MMICs using SiGe HBTs," *IEICE Trans. Electronics*, vol. 82-C, pp. 1943-1950, 1999.

[71] K.I. Jeon, Y.S. Kwon, and S.C. Hong, "Input harmonics control using non-linear capacitor in GaAs FET power amplifier," *IEEE MTT-S International Microwave Symposium Digest*, pp. 817-820, Jun. 1997.

[72] P.J.v. Wijnen, H.R. Claessen, and E.A. Wolsheimer, "A new straightforward calibration and correction procedure for "on wafer" high-frequency s-parameter measurements (45 MHz-18 GHz)," *Proceedings of IEEE Bipolar/BiCMOS Circuits and Technology Meeting*, pp. 70-73, 1987.

[73] H. Cho and D.E. Burk, "A three-step method for the de-embedding of high-frequency s-parameter measurements," *IEEE Trans. Electron Devices*, vol. 38, pp. 1371-1375, June 1991.

[74] J. Grzyb, D. Cottet, and G Tröster, "Systematic de-embedding of the transmission line parameters on high-density substrates with probe-tip calibrations," *IEEE Electronic Components and Technology Conference*, pp. 1051-1057, 2002.

[75] R.A. Pucel, W. Struble, R. Hallgren, and U.L. Rohde, "A general noise de-embedding procedure for packaged two-port linear active devices," *IEEE Trans. Microwave Theory and Tech.*, vol. 40, pp. 2013-2024, Nov. 1992.

[76] Q. Liang, J.D. Cressler, G. Niu, Y. Lu, G. Freeman, D.C. Ahlgren, R.M. Malladi, K. Newton, and D.L. Harame. "A simple four-port parasitic de-embedding methodology for high-frequency scattering parameter and noise characterization of SiGe HBTs," *IEEE Trans. Microwave Theory and Tech.*, vol. 51, pp. 2165-2174, Nov. 2003.

[77] V. Rizzoli, F. Mastri, and C. Cecchetti, "Computer-aided noise analysis of MESFET and HEMT mixers," *IEEE Trans. Microwave Theory and Tech.*, vol. 37, pp. 1401-1410, Sept. 1989.

[78] S. Lee, V. Tilak, K.J. Webb, and L.F. Eastman, "Intrinsic noise characteristics of AlGaN/GaN HEMTs," *IEEE MTT-S International Microwave Symposium Digest*, pp. 1415-1418, 2002.

[79] T. Blalack, Y. Leclercq, and C.P. Yue, "On-chip RF isolation techniques," *Proceedings of IEEE Bipolar/BiCMOS Circuits and Technology Meeting*, pp. 205-211, 2002.

[80] R. Gharpurey and R.G. Meyer, "Modeling and analysis of substrate coupling in integrated circuits," *IEEE J. Solid-State Circuits*, vol. 31, pp. 344-353, Mar. 1996.

[81] K.L. Fong and R.G. Meyer. "High-frequency nonlinearity analysis of common-emitter and differential-pair transconductance stages," *IEEE J. Solid-State Circuits*, vol. 33, pp. 548-555, 1998.

VITA

Qingqing Liang was born in Changsha, Hunan, P.R. China, in 1977. He earned his B.S. and M.S. degrees in electrical engineering from Fudan University, Shanghai, China, in 1997 and 2000, respectively, and is currently working toward a Ph.D. degree in electrical and computer engineering in Georgia Institute of Technology, Atlanta, GA.

His research focuses on the modeling, simulation and characterization of SiGe HBT devices and circuits, design and optimization of RF building-blocks at both the device and circuit level.